Making Sense of the SOCIAL WORLD

METHODS OF INVESTIGATION

5

Daniel F. Chambliss
Russell K. Schutt

Making Sense of the Social World

Fifth Edition

Related Titles in Research Methods and Statistics From SAGE

Investigating the Social World: The Process and Practice of Research, Eighth Edition, by Russell K. Schutt

The Practice of Research in Social Work, Third Edition, by Rafael J. Engel and Russell K. Schutt

The Practice of Research in Criminology and Criminal Justice, Fifth Edition, by Ronet K. Bachman and Russell K. Schutt

Fundamentals of Research in Criminology and Criminal Justice, Third Edition, by Ronet K. Bachman and Russell K. Schutt

Designing Surveys: A Guide to Decisions and Procedures, Second Edition, by Ronald F. Czaja and Johnny Blair

A Guide to Qualitative Field Research, Second Edition, by Carol A. Bailey

Adventures in Social Research, Seventh Edition, by Earl Babbie, Fred Halley, William E. Wagner, III, and Jeanne Zaino

Adventures in Criminal Justice Research, Fourth Edition, by Kim Logio, George Dowdall, Earl Babbie, and Fred Halley

Social Statistics for a Diverse Society, Sixth Edition, by Chava Frankfort-Nachmias and Anna Leon-Guerrero

Essentials of Social Statistics for a Diverse Society by Anna Leon-Guerrero and Chava Frankfort-Nachmias

Research Methods in Psychology, Second Edition, by Paul G. Nestor and Russell K. Schutt

Research Methods in Education by Joseph Check and Russell K. Schutt

Fundamentals of Social Work Research, Second Edition, by Rafael J. Engel and Russell K. Schutt

Constructing Social Research by Charles C. Ragin and Lisa M. Amoroso

Using IBM® SPSS® Statistics for Social Statistics and Research Methods, Third Edition, by William E. Wagner, III

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore | Boston

Making Sense of the Social World

Methods of Investigation

Fifth Edition

Daniel F. Chambliss
Hamilton College
Russell K. Schutt
University of Massachusetts Boston

FOR INFORMATION:

SAGE Publications, Inc.

2455 Teller Road

Thousand Oaks, California 91320

E-mail: order@sagepub.com

SAGE Publications Ltd.

1 Oliver’s Yard

55 City Road

London EC1Y 1SP

United Kingdom

SAGE Publications India Pvt. Ltd.

B 1/I 1 Mohan Cooperative Industrial Area

Mathura Road, New Delhi 110 044

India

SAGE Publications Asia-Pacific Pte. Ltd.

3 Church Street

#10-04 Samsung Hub

Singapore 049483

Copyright © 2016 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Printed in the United States of America

A catalog record of this book is available from the Library of Congress.

ISBN 978-1-4833-8061-2 (pbk.)

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Jerry Westby

Associate Editor: Jessica Miller

Associate Digital Content Editor: Nick Pachelli

Editorial Assistant: Laura Kirkhuff

Production Editor: David C. Felts

Copy Editor: Robin Gold

Typesetter: C&M Digitals (P) Ltd.

Proofreader: Tricia Currie-Knight

Indexer: Kathy Paparchontis

Cover Designer: Gail Buschman

Marketing Manager: Erica DeLuca

Detailed Contents

Detailed Contents

- About the Authors
- Preface
- Acknowledgments
- Chapter 1. Science, Society, and Social Research
  - Learning About the Social World
    - Overgeneralization
    - Selective or Inaccurate Observation
    - Illogical Reasoning
    - Resistance to Change
  - Research That Matters
  - Can Social Scientists See the Social World More Clearly?
    - Social Research in Practice
      - Description: How Often Do Americans “Neighbor”?
      - Exploration: How Do Athletic Teams Build Player Loyalty?
  - Research in the News: Why Doesn’t the Internet Reach Everyone?
    - Explanation: Does Social Context Influence Adolescent Outcomes?
    - Evaluation: Does More Social Capital Result in More Community Participation?
  - Careers and Research: Jessica LeBlanc, Research Assistant
  - How Well Have We Done Our Research?
    - Measurement Validity
    - Generalizability
    - Causal Validity
  - Conclusion
  - Key Terms
  - Highlights
  - STUDENT Study Site
  - Exercises
- Chapter 2. The Process and Problems of Social Research
  - What Is the Question?
  - What Is the Theory?
  - What Is the Strategy?
    - Deductive Research
  - Research in the News: Investigating Child Abuse Doesn’t Reduce It

Domestic Violence and the Research Circle
Inductive Research
Research That Matters
An Inductive Study of Response to a Disaster
Descriptive Research: A Necessary Step
What Is the Design?
Cross-Sectional Designs
Longitudinal Designs
Trend Designs
Panel Designs
Cohort Designs
Units and Levels of Analysis
Careers and Research: Russell K. Schutt, PhD
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 3. Ethics in Research
Historical Background
Ethical Principles
Protecting Research Subjects
Avoid Harming Research Participants
Obtain Informed Consent
Avoid Deception in Research, Except in Limited Circumstances
Maintain Privacy and Confidentiality
Research That Matters
Maintaining Honesty and Openness
Research in the News: What Would an IRB Say?
Achieving Valid Results
Careers and Research: Kristen Kenny, Research Compliance Specialist
Encouraging Appropriate Application
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 4. Conceptualization and Measurement

- What Do We Have in Mind?
  - Conceptualization
- Research That Matters
- Research in the News: What Is Your Race?
  - Variables and Constants
- How Will We Know When We've Found It?
  - Using Available Data
  - Content Analysis
  - Constructing Questions
    - Single Questions
    - Indexes and Scales
  - Making Observations
  - Combining Measurement Operations
- How Much Information Do We Really Have?
  - Nominal Level of Measurement
  - Ordinal Level of Measurement
  - Interval Level of Measurement
  - Ratio Level of Measurement
  - Comparison of Levels of Measurement
- Careers and Research: Dana Hunt, Principal Scientist
- Did We Measure What We Wanted to Measure?
  - Measurement Validity
    - Face Validity
    - Criterion Validity
    - Construct Validity
  - Reliability
  - Can We Achieve Both Reliability and Validity?
- Conclusion
- Key Terms
- Highlights
- STUDENT Study Site
- Exercises

Chapter 5. Sampling and Generalizability

- How Do We Prepare to Sample?
  - Define Sample Components and the Population
  - Evaluate Generalizability
  - Assess the Diversity of the Population
  - Consider a Census
- Research That Matters

- What Sampling Method Should We Use?
  - Probability Sampling Methods
- Research in the News: What Are Best Practices for Sampling Vulnerable Populations?
  - Simple Random Sampling
  - Systematic Random Sampling
  - Cluster Sampling
  - Stratified Random Sampling
- Nonprobability Sampling Methods
  - Availability Sampling
- Careers and Research: Ross Koppel, Sociologist
  - Quota Sampling
  - Purposive Sampling
  - Snowball Sampling
- Conclusion
- Key Terms
- Highlights
- STUDENT Study Site
- Exercises

Chapter 6. Causation and Experimental Design

- Causal Explanation
- What Causes What?
  - Association
- Research in the News: Where Did You Hear That?
  - Time Order
  - Nonspuriousness
- Research That Matters
  - Mechanism
  - Context
- Why Experiment?
- Careers and Research: Sruthi Chandrasekaran, Senior Research Associate
- What If a True Experiment Isn’t Possible?
  - Nonequivalent Control Group Designs
  - Before-and-After Designs
  - Ex Post Facto Control Group Designs
- What Are the Threats to Validity?
  - Threats to Internal (Causal) Validity
    - Noncomparable Groups

Endogenous Change
History
Contamination
Treatment Misidentification
Threats to Generalizability
Sample Generalizability
Cross-Population Generalizability
Interaction of Testing and Treatment
How Do Experimenters Protect Their Subjects?
Deception
Selective Distribution of Benefits
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 7. Survey Research
Why Is Survey Research So Popular?
Research in the News: What Can Surveys Uncover?
How Should We Write Survey Questions?
Be Clear; Avoid Confusing Phrasing
Minimize Bias
Allow for Disagreement
Don’t Ask Questions They Can’t Answer
Allow for Uncertainty
Make Response Categories Exhaustive and Mutually Exclusive
How Should Questionnaires Be Designed?
Build on Existing Instruments
Refine and Test Questions
Maintain Consistent Focus
Order the Questions
Careers and Research: Floyd J. (“Jack”) Fowler Jr., Founder and Director of the Center for Survey Research
Make the Questionnaire Attractive
What Are the Alternatives for Administering Surveys?
Mailed, Self-Administered Surveys
Group-Administered Surveys
Telephone Surveys
Reaching Sampling Units

- In-Person Interviews
- Research That Matters
- Maximizing Response to Interviews
  - Electronic Surveys
- A Comparison of Survey Designs
- Ethical Issues in Survey Research
- Conclusion
- Key Terms
- Highlights
- STUDENT Study Site
- Exercises

Chapter 8. Elementary Quantitative Data Analysis

- Why Do Statistics?
  - Case Study: The Likelihood of Voting
- How to Prepare Data for Analysis
- What Are the Options for Displaying Distributions?
  - Graphs
  - Frequency Distributions
- What Are the Options for Summarizing Distributions?
- Research in the News: General Social Survey Shows Infidelity on the Rise
  - Measures of Central Tendency
    - Mode
    - Median
    - Mean
    - Median or Mean?
  - Measures of Variation
- Research That Matters
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation
- How Can We Tell Whether Two Variables Are Related?
  - Reading the Table
  - Controlling for a Third Variable
- Secondary Data Analysis
- Careers and Research: Claire Wulf Winiarek, Director of Collaborative Policy Engagement
  - U.S. Census Bureau

Bureau of Labor Statistics (BLS)
Inter-university Consortium for Political and Social Research
Human Relations Area Files
Big Data
Ethical Issues in Statistical Analysis, Secondary Data Analysis, and Big Data
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 9. Qualitative Methods: Observing, Participating, Listening
What Are Qualitative Methods?
Case Study: Beyond Caring
Ethnography
Careers and Research: Amanda Aykanian, Research Associate
Netnography
Research in the News: Can Taping Interviews Capture a Trend?
Ethnomethodology
How Does Participant Observation Become a Research Method?
Choosing a Role
Complete Observation
Mixed Participation or Observation
Complete Participation
Entering the Field
Developing and Maintaining Relationships
Research That Matters
Sampling People and Events
Taking Notes
Managing the Personal Dimensions
Systematic Observation
How Do You Conduct Intensive Interviews?
Establishing and Maintaining a Partnership
Asking Questions and Recording Answers
Interviewing Online
How Do You Run Focus Groups?
Ethical Issues in Qualitative Research
Conclusion
Key Terms

Highlights
STUDENT Study Site
Exercises
Chapter 10. Qualitative Data Analysis
What Is Distinctive About Qualitative Data Analysis?
Qualitative Data Analysis as an Art
Qualitative Compared With Quantitative Data Analysis
What Techniques Do Qualitative Data Analysts Use?
Documentation
Conceptualization, Coding, and Categorizing
Examining Relationships and Displaying Data
Authenticating Conclusions
Research That Matters
Reflexivity
What Are Some Alternatives in Qualitative Data Analysis?
Narrative Analysis
Conversation Analysis
Grounded Theory
Careers and Research: Laurel Person Mecca, Assistant Director and Senior Research Specialist
Visual Sociology
Why Are Mixed Methods Helpful?
Research in the News: What's in a Message?
How Can Computers Assist Qualitative Data Analysis?
What Ethical Issues Arise in Doing Qualitative Data Analysis?
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 11. Unobtrusive Measures
Creative Sources
Physical Traces
Archives
Observation
Contrived Observation
Content Analysis
Identify a Population of Documents or Other Textual Sources
Determine the Units of Analysis

Select a Sample of Units From the Population
Design Coding Procedures for the Variables to Be Measured
Develop Appropriate Statistical Analyses
Historical Methods
Event-Structure Analysis
Oral History
Comparative Methods
Careers and Research: Kurt Taylor Gaubatz, Faculty, Department of Political Science & Geography International Studies Graduate Program
Cross-Sectional Comparative Research
Research in the News: Can We Compare Equality on the Board Across the Globe?
Longitudinal Comparative Research
Research That Matters
Cautions for Comparative Analysis
Ethical Issues in Unobtrusive Methods
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 12. Evaluation Research
What Is the History of Evaluation Research?
Careers and Research: Mary Anne Casey, Consultant
What Is Evaluation Research?
What Are the Alternatives in Evaluation Designs?
Black Box or Program Theory
Researcher or Stakeholder Orientation
Quantitative or Qualitative Methods
Simple or Complex Outcomes
What Can an Evaluation Study Focus On?
Needs Assessment
Research in the News: What Motivates Policy Shifts?
Evaluability Assessment
Process Evaluation
Research That Matters
Impact Analysis
Efficiency Analysis

Ethical Issues in Evaluation Research
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Chapter 13. Reviewing, Proposing, and Reporting Research
Comparing Research Designs
Reviewing Research
Case Study: “Night as Frontier”
The Research Design
Analysis of the Design
An Overall Assessment
Case Study: When Does Arrest Matter?
Summarize Prior Research
Critique Prior Research
Present Pertinent Conclusions
Research That Matters
Proposing New Research
Case Study: Community Health Workers and Cancer Clinical Trials
Careers and Research: Ruth Westby, Research Associate
Reporting Research
Research in the News: How Much Should Social Scientists Report?
Writing and Organizing
Plagiarism
Conclusion
Key Terms
Highlights
STUDENT Study Site
Exercises
Glossary
Appendix A: Finding Information
References
Index

About the Authors

Daniel F. Chambliss,

PhD, is the Eugene M. Tobin Distinguished Professor of Sociology at Hamilton College in Clinton, New York, where he has taught since 1981. He received his PhD from Yale University in 1982; later that year, his thesis research received the American Sociological Association's Medical Sociology Dissertation Prize. In 1988, he published the book *Champions: The Making of Olympic Swimmers*, which received the Book of the Year Prize from the U.S. Olympic Committee. In 1989, he received the American Sociology Association (ASA)'s Theory Prize for work on organizational excellence based on his swimming research. Recipient of both Fulbright and Rockefeller Foundation fellowships, he published his second book, *Beyond Caring: Hospitals, Nurses, and the Social Organization of Ethics*, in 1996; for that work, he was awarded the ASA's Elliot Freidson Prize in Medical Sociology. In 2014, Harvard University Press published his book, *How College Works*, coauthored with his former student Christopher G. Takacs. His research and teaching interests include organizational analysis, higher education, social theory, and comparative research methods.

Russell K. Schutt,

PhD, is professor and chair of sociology at the University of Massachusetts Boston, where he received the 2007 Chancellor's Award for Distinguished Service. Since 1990, he has also been lecturer on sociology in the Department of Psychiatry (Beth Israel Deaconess Medical Center) at the Harvard Medical School. He completed his BA, MA, and PhD degrees at the University of Illinois at Chicago and was a postdoctoral fellow in the Sociology of Social Control Training Program at Yale University. In addition to eight editions of the text on which this brief edition is based, *Investigating the Social World: The Process and Practice of Research*, and four other coauthored versions—for the fields of social work, criminal justice, psychology, and education—his other books include *Homelessness, Housing, and Mental Illness* (2011),*Social Neuroscience: Brain, Mind, and Society* (coedited, 2015), *and Organization in a Changing Environment* (1986). He has authored and coauthored numerous journal articles, book

chapters, and research reports on homelessness, mental health, organizations, law, and teaching research methods. His research has included a mixed-methods investigation of a public health coordinated care program, a study of community health workers and recruitment for cancer clinical trials, a mixed-methods study of a youth violence reduction program, a randomized trial of a peer support program for homeless dually diagnosed veterans, and a randomized evaluation of housing alternatives for homeless persons diagnosed with severe mental illness, with funding from the National Cancer Institute, the Veterans Health Administration, the National Institute of Mental Health, the John E. Fetzer Institute, and state agencies. His current scholarly foci are the impact of the social environment on cognitive and community functioning, the meaning of housing and service preferences, and the value of alternative organizational and occupational structures for service delivery. His prior research has also included investigation of social factors in legal decisions and admission practices and of influences on job and service satisfaction. Details are available at http://rschutt.wikispaces.umb.edu.

Preface

If you have been eager to begin your first course in social science research methods, we are happy to affirm that you've come to the right place. We have written this book to give you just what you were hoping for—an introduction to research that is interesting, thoughtful, and thorough.

But what if you've been looking toward this course with dread, putting it off for longer than you should, wondering why all this "scientific" stuff is required of students who are really seeking something quite different in their major? Well, even if you had just some of these thoughts, we want you to know that we've had your concerns in mind, too. In *Making Sense of the Social World*, we introduce social research with a book that combines professional sophistication with unparalleled accessibility: Any college student will be able to read and understand it—even enjoy it—and experienced social science researchers, we hope, can learn from our integrated approach to the fundamentals. And whatever your predisposition to research methods, we think you'll soon realize that understanding them is critical to being an informed citizen in our complex, fast-paced social world.

Teaching and Learning Goals

Our book will introduce you to social science research methods that can be used to study diverse social processes and to improve our understanding of social issues. Each chapter illustrates important principles and techniques in research methods with interesting examples drawn from formal social science investigations and everyday experiences.

Even if you never conduct a formal social science investigation after you complete this course, you will find that improved understanding of research methods will sharpen your critical faculties. You will become a more informed consumer, and thus a better user, of the results of the many social science studies that shape social policy and popular beliefs. Throughout this book, you will learn what questions to ask when critiquing a research study and how to evaluate the answers. You can begin to sharpen your critical teeth on the illustrative studies throughout the book. Exercises at the end of each chapter will allow you to find, discuss, critique, and actually do similar research.

If you are already charting a course toward a social science career, or if you decide to do so after completing this course, we aim to give you enough “how to” instruction so that you can design your own research projects. We also offer “doing” exercises at the end of each chapter that will help you try out particular steps in the research process.

But our goal is not just to turn you into a more effective research critic or a good research technician. We do not believe that research methods can be learned by rote or applied mechanically. Thus, you will learn the benefits and liabilities of each major research approach as well as the rationale for using a combination of methods in some situations. You will also come to appreciate why the results of particular research studies must be interpreted within the context of prior research and through the lens of social theory.

Organization of the Book

The first three chapters introduce the why and how of research in general. Chapter 1 shows how research has helped us understand how social relations have changed in recent years and the impact of these changes. Chapter 2 illustrates the basic stages of research with studies of domestic violence, Olympic swimmers, and environmental disasters. Chapter 3 introduces the ethical considerations that should guide your decisions throughout the research process. The next three chapters discuss how to evaluate the way researchers design their measures (Chapter 4), draw their samples (Chapter 5), and justify their statements about causal connections (Chapter 6).

As we present the logic of testing causal connections in Chapter 6, we also present the basics of the experimental designs that provide the strongest tests for causality. In Chapter 7, we cover the most common method of data collection in sociology—surveys—and in Chapter 8, we present the basic statistical methods that are used to analyze the results of the quantitative data that often are collected in experiments and surveys. Here we examine the results of the 2010 General Social Survey to see how these statistics are used.

Chapters 9, 10, and 11 shift the focus from strategies for collecting and analyzing quantitative data to strategies for collecting and analyzing qualitative data. In Chapter 9, we focus on the basic methods of collecting qualitative data: participant observation and ethnography, intensive interviews, and focus groups. We also introduce approaches such as ethnomethodology and netnography. In Chapter 10, we review the logic of qualitative data analysis and several specific approaches: narrative analysis, conversation analysis, and grounded theory, as well as the "mixed methods" approach that combines various methods. In Chapter 11, we introduce the array of "non-obtrusive measures" in which the process of research will not in itself change what is being studied—it's nonreactive. Chapter 12 explains how you can combine different methods to evaluate social programs. Chapter 13 covers the review of prior research, the development of research proposals, and the writing and reporting of research results.

Distinctive Features of This Edition

In making changes for this edition, we feel we have advanced even further in pursuit of our goal of making research methods one of your most enjoyable and engaging courses. We have incorporated valuable suggestions from many faculty reviewers and students who have used the book over the several years since it was first released. As in the previous four editions, this book has also benefited from advances in its parent volume, Russell Schutt's *Investigating the Social World: The Process and Practice of Research* (now in its eighth edition).

A new chapter on unobtrusive measures.

Prompted by reviewers, and encouraged by reader responses to some examples in the previous edition, we have expanded our coverage of nonreactive measures, with many new examples of "creative" sources of data, as well as integrating content analysis, historical, and comparative methods.

New material on the role of social media in research.

To reflect recent developments in social media, we have incorporated a variety of examples of research based on digital technologies and extended discussions of how such media modify research techniques both in survey and in qualitative methods.

A major expansion of coverage of mixed methods.

At the suggestion of our reviewers, we have expanded the coverage of macro forms of research and demonstrated the advantages of mixing methods to balance the strengths and weaknesses of different approaches.

Expanded coverage of quantitative methods.

In Chapter 8 ("Elementary Quantitative Data Analysis"), a majority of the examples and tables have been updated, and major new sections have been added on secondary data analysis and big data.

Major clarification of difficult or especially important topics.

Again prompted by reviewers, we have rewritten and clarified sections on various topics: variables, operationalization, types of reliability, causal mechanisms, and context.

Updated information on the impact of cell phones and the web on survey response.

Chapter 7 on survey research provides extensive recent information on the impact of the increasing use of cell phones and web surveys on survey practice.

Careers and Research features.

These features have been added to each chapter to help students see how what they learn can actually be used in today's job markets.

Research That Matters features.

These features show how important social research figures in many of the major issues of the day.

Our text also offers other distinctive features:

Brief examples of social research.

In each chapter, these illustrate particular points and show how research techniques are used to answer important social questions. Whatever your particular substantive interests in social science, you'll find some interesting studies that will arouse your curiosity.

Integrated treatment of causality and experimental design.

We have combined the discussions of causation and experimental design in order to focus on the issues that are most often encountered during research in sociology, criminal justice, education, social work, communications, and political science.

Realistic coverage of ethical concerns and ethical decision making.

Like the parent volume, *Investigating the Social World*, this text presents ethical issues that arise in the course of using each method of data collection, as well as comprehensive coverage of research ethics in a new chapter.

Engaging end-of-chapter exercises.

We organize the exercises under the headings of discussing, finding, critiquing, and doing, and end with questions about ethics. New exercises have been added, and some of the old ones have been omitted. The result is a set of learning opportunities that should greatly facilitate the learning process.

Software-based learning opportunities.

The text's website (edge.sagepub.com/chamblissmssw5e) includes review exercises to help you master the concepts of social research, a set of articles that provide examples of different methods, and a portion of the 2010 General Social Survey (GSS) so you can try out quantitative data analysis (if your school provides access to the SPSS statistical package).

Aids to effective study.

Lists of main points and key terms provide quick summaries at the end of each chapter. In addition, key terms are highlighted in boldface type when first introduced and defined in the text. Definitions of key terms can also be found in the glossary/index at the end of the book. The text's website (edge.sagepub.com/chamblissmssw5e) offers more review questions. An instructor's manual includes more exercises that have been specially designed for collaborative group work in and outside of class. Appendix A, Finding Information, provides up-to-date information about using the Internet.

In the electronic edition of the book you have purchased, there are several icons that reference links (videos, journal articles) to additional content. Though the electronic edition links are not live, all content referenced may be accessed at . This URL is referenced at several points throughout your electronic edition.

Acknowledgments

First, we would like to thank Jerry Westby, publisher and senior editor at SAGE, our main managerial contact and source of encouragement as we developed our text. Other members of the SAGE team helped in multiple ways: Megan Krattli provided expert assistance with the review of the reviews, David Felts smoothly guided the manuscript through the production process, and Robin Gold did a fine job of copyediting.

The reviewers for this edition helped us to realize the potential for the revision. We are very grateful for the wisdom and critical acumen of the following:

Vandeen Campbell, Lehman College
Amy Dennissen, University of Colorado
Candan Duran-Aydintug, University of Colorado Denver
Martie Gillen, University of Florida
Vanessa P. Jackson, University of Kentucky
Jill Littrell, PhD, Georgia State University
Liz Marie Marciniak, University of Pittsburgh at Greensburg
Deborah A. Potter, PhD, University of Louisville
Linda Rouse, University of Texas
Fritz W. Yarrison, Kent State University

And for previous editions:

Sandy D. Alvarez, Indiana State University
James David Ballard, California State University, Northridge
Diane C. Bates, Sam Houston State University
Irenee R. Beattie, University of California, Merced
Julio Borquez, University of Michigan–Dearborn
Rebecca Brooks, Ohio Northern University
Matthew W. Brosi, Oklahoma State University
Matthew M. Caverly, University of Florida
Jin Young Choi, Sam Houston State University
Cristina Bodinger-deUriarte, California State University, Los Angeles
Keith F. Durkin, Ohio Northern University
Mark Edwards, Oregon State University

Juanita M. Firestone, The University of Texas at San Antonio
David Folz, The University of Tennessee, Knoxville
Kellie J. Hagewen, University of Nebraska–Lincoln
Dena Hartley, University of Akron
Laura Hecht, California State University, Bakersfield
Jerome L. Himmelstein, Amherst College
Margaret Platt Jendrek, Miami University (Oxford)
Julie A. Kmec, Washington State University
Ann Marie Kinnell, The University of Southern Mississippi
Manfred Kuechler, Hunter College
Vera Lopez, Arizona State University
Karen McCue, University of New Mexico
Dianne Mosley, Texas Southern University
Ed Nelson, California State University, Fresno
Sookhee Oh, University of Missouri–Kansas City
Colin Olson, University of New Mexico
Kate Peirce, Texas State University
Ronald Perry, Arizona State University
Travis N. Ridout, Washington State University
Nick Sanyal, University of Idaho
Steve Swinford, Montana State University
Felicia P. Wiltz, Suffolk University
Chenyang Xiao, Washington State University
Xiaohe Xu, The University of Texas at San Antonio
David Zefir, Plymouth State College
Kristen Zgoba, Rutgers University

We are grateful to Peter Marsden, of Harvard University, who updated us on web surveys; to Philip Brenner, of the University of Massachusetts Boston, who informed us about nonreponse error; and to Sunshine Hillygus, former director of Harvard University's Program on Survey Research, for sharing with us findings about survey response rates. Steve Ellingson and Bonnie Urciouli of Hamilton College made valuable suggestions, as did Morten Ender of the U.S. Military Academy and Dan Ryan Jr. of Mills College and Elizabeth Schneider (Russ's wife) with respect to Appendix A. We thank Whitney Gecker and Kate Russell for writing tracks for the interactive exercises that link to SAGE journal articles, Megan Reynolds and Kathryn Stoeckert for their work on previous sets of exercises, Amanda Colligan and Robyn Brown for their help on the eBook, and Vital Source (and the former VPG Integrated Media) for the online

programming.

We also have some personal thank-yous:

Dan Chambliss: I wish gratefully to acknowledge the assistance, in many areas, of Marcia Wilkinson, who as typist, transcriber, organizer, and administrative aide and daily conscience is simply irreplaceable. My students at Hamilton College have been a blessing throughout: Chris Takacs helped to design and create several of the exhibits, solving intellectual problems through graphic displays; Shauna Sweet told me where the book was good and where it wasn't, clarified the regression effect, and showed me how people actually should read contingency tables; Katey Healy-Wurzburg, in one of many moments of intellectual brilliance, explained the underlying rhetorical problem of the ecological fallacy; and Erin Voyik, as a teaching assistant in my Methods class, laid out for me time and again what students do and don't actually understand, and enjoy, about social research methods. Many others, students and colleagues alike, have contributed without recognition; let's just say that all of intellectual life is communal, and we fully appreciate that fact. And finally, I hope that my wife, Susan Morgan, enjoyed, at least vicariously, the thrills I felt in working on this book as much as I enjoyed sharing them with her.

Russ Schutt: I am grateful to the many reviewers of this and previous editions of *Making Sense of the Social World* and of *Investigating the Social World*, as well as to the many staff and consultants at SAGE who helped make these texts a success. My special thanks to Whitney Gecker, the outstanding graduate student who checked web exercises and updated statistical examples throughout the text and wrote new interactive exercises based on the "Research That Matters" highlights in this edition. I also want to express my appreciation to my many research collaborators, with whom I have shared so many fascinating and educational experiences and from whom I have learned so much, and for the many fine students at the University of Massachusetts, Boston, who continue to renew my enthusiasm for quality teaching. Most importantly, I thank my wife, Elizabeth Schneider, for her ongoing support and love, and my daughter, Julia, for being such a marvelous young woman.

xx**Dan and Russ:** Finally, Dan wants to say that Russ Schutt is a wonderful coauthor, with whom he is honored to work: totally responsible and respectful, hardworking, serious in his scholarship but without a trace of arrogance. His generous personality has allowed this collaboration to sail along beautifully.

Russ adds that Dan is the perfect model of the gentleman and scholar, whose research savvy and keen intelligence are matched to a warm and caring persona. We both like to think that our talents are almost entirely complementary. We are immensely grateful for the chance to work together.

Dan:

To my sweetheart Susan and to the gifts she brought to me—

Sarah, Daniel, Anne, and Rebecca

Russ:

To Beth and Julia

Chapter 1 Science, Society, and Social Research

©iStockphoto.com/mattjeacock

Learning Objectives

1. Describe the four common errors in everyday reasoning.
2. Define *social science* and identify its limitations.
3. Identify the four goals for social research in practice.
4. Define valid knowledge and indicate the three components of validity.

Online networking services added a new dimension to social life in the early years of the 21st century. Mark Zuckerberg started Facebook in 2004 as a service for college students like himself, but by September 30, 2013, Facebook (2013) had grown to be a global service with more than 1.19 billion users—more than one of every six people in the world and four out of every five persons in the United States (Internet World Stats 2012; Statistic Brain 2013; U.S. Census Bureau 2013).

Has social networking helped you keep in touch with your friends, or make new friends? Has it changed your face-to-face interactions with other people? Has it improved, or damaged, your social life? That's where social researchers begin, with questions about the social world and a desire to find the answers. Please answer the following:

1. Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?
   __ Most people can be trusted
   __ You can't be too careful
2. Do you use the Internet, at least occasionally?
   __ Use Internet
   __ Do not use Internet
3. Did you happen to use the Internet YESTERDAY?
   __ Yes
   __ No
4. Counting all of your online sessions, how much time did you spend using the Internet yesterday?
   __ Less than 15 minutes
   __ 15 minutes to less than half hour

__ Half hour or more but less than an hour
__ More than 1 hour but less than 2 hours
__ 2 hours or more but less than 3 hours
__ 3 hours or more but less than 4 hours
__ 4 hours or more

5. Please tell me if you ever use the Internet to do any of the following things. Do you ever use a social networking site like Myspace, Facebook, or LinkedIn?
   __ Yes
   __ No
6. Have you made a friend or contact on a social networking website like Myspace, Facebook, or LinkedIn?
   __ Yes
   __ No
7. Do you belong to or ever work with a community group or neighborhood association that focuses on issues or problems in your community?
   __ Yes
   __ No

These are questions from the Social Networking Sites and Facebook Survey conducted by Princeton Survey Research Associates International for the Pew Research Center Internet & American Life Project. When Keith N. Hampton, Lauren Sessions Goulet, Lee Rainie, and Kirsten Purcell (2011) analyzed the responses, they found that 79% of U.S. adults ages 18 to 22 use the Internet, 59% use social networking services, and that networking usually supports people's other social connections, rather than displacing them.

Video Link

Look at social research findings on the use of social media.

Social research such as this differs from everyday thinking by asking broader questions, about people outside our immediate experience, about why (not just what) things happen, and by using systematic research methods. In this chapter, we hope to convince you that careful social research produces knowledge that can be more important, more trustworthy, and more useful than can personal opinions or individual experiences. By the chapter's end, you should know what is "scientific" in social science and appreciate how the methods of science can

help us understand the problems of society.

Learning About the Social World

Consider some questions that social scientists have asked about the Internet and social ties:

1. What percentage of Americans are connected to the Internet?

The 2011 Current Population Survey by the U.S. Census Bureau (File 2013a) of approximately 54,000 households revealed that 75% of U.S. households had a computer at home and almost that many were connected to the Internet (File 2013a). The Pew Research Center's Internet & American Life Project 2010 survey of 2,255 adult Americans found that almost half used a social network service (Hampton et al. 2011). These percentages have increased rapidly since personal computers first came into use in the early 1980s and after the Internet became publicly available in the 1990s.

2. How does Internet use vary across social groups?

Internet use differs dramatically between social groups. Question #2 in the little survey earlier ("Do you use the Internet . . . ?") may have seemed obvious to you—of course you do!—but you're a college student. It turns out that other people don't use the Internet nearly as much. As indicated in Exhibit 1.1, Internet use in 2011 ranged from as low as 32% among those with less than a high school education to 90% among those with at least a bachelor's degree (File 2013a). People with little education are quite likely never to use the Internet, even though use has certainly increased for all education levels since 1997 (Strickling 2010). Internet use also is greater with higher family income and is higher among whites and Asian Americans than among Hispanic Americans and black Americans (Cooper & Gallager 2004: Appendix, Table 1). Social network sites are most heavily used by younger people (89% of those age 18 to 29), compared with those who are middle aged (78% among those age 30 to 49 and 60% of those age 50 to 64) or senior citizens (43% of those age 65 and older) (Pew Research Center 2013).

3. Does Internet use interfere with maintaining other social ties?

It doesn't seem so. The rate of social isolation—that is, of people not having anyone to confide in—only rose from 8% in 1985 to 12% in 2008, during the

Internet's boom years. According to the Pew Center survey, individuals who use the Internet actually tend to have larger and more diverse social networks and are equally as likely as are non-users to participate in community activities. However, a different survey back in 2004 led other researchers to conclude that social isolation had increased considerably since 1985 (Marsden 1987; McPherson, Smith-Lovin, & Brashears 2006: 358). Even careful social science research projects can disagree like this, so we learn in this book how to evaluate their methods and findings.

4. Does wireless access (Wi-Fi) in such public places as Starbucks decrease social interaction among customers?

It depends on the customer. Keith Hampton and Neeti Gupta (2008) observed Internet use in coffee shops with wireless access in two cities and concluded that there were two types of Wi-Fi users: Some used their Internet connection to create a secondary work office, and others used their Internet connection as a tool for meeting others in the coffee shop. This means that Wi-Fi was associated with less interaction for some customers, but more for others. (See "Research That Matters," later in this chapter).

5. Do cell phones and e-mail hinder the development of strong social ties?

Again, it seems to depend. Based on surveys in Norway and Denmark, Rich Ling and Gitte Stald (2010) concluded that mobile phones increase social ties among close friends and family members, whereas e-mail communication tends to decrease ties, but research by the Pew Center has identified positive effects of the Internet and e-mail on social ties (Boase et al. 2006).

Did you expect these results? You have seen that people with more education use the Internet more than do those with less education. Did you know the difference was so large (#2)? Maybe you have heard people gripe about the effect of the Internet on relationships. Is it safe to draw general conclusions from such complaints (#3)? Have you noticed the effects of surroundings and mode of communication on different people (#4 and #5)?

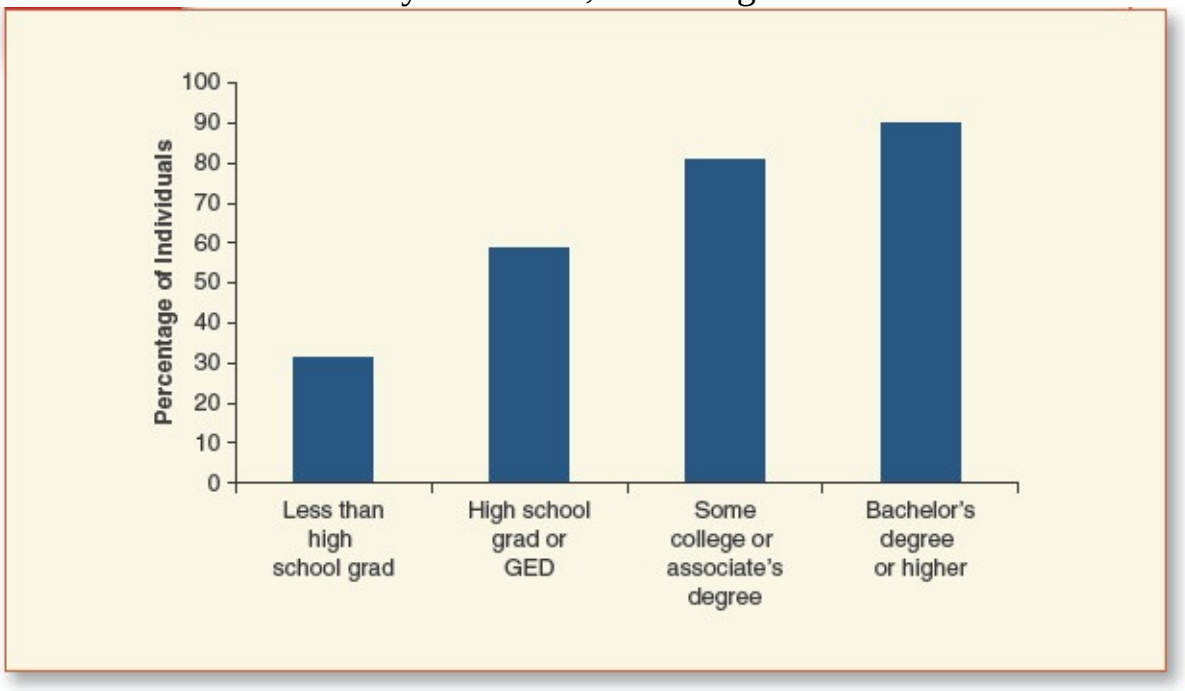
The more that you begin to "think like a social scientist," the more such questions will come to mind. But as you've just seen, in our everyday reasoning about the social world, our own prior experiences and orientations can have a major influence on how we perceive and interpret the world. As a result, one

person may think that posting Facebook updates is “what’s wrong with modern society,” but another might see the same behavior as helping people to “get connected.” Social science is an effort to get beyond personal biases and limited viewpoints to something at least a bit more objective.

Audio Link

Listen to how online relationships impacts social research.

Exhibit 1.1 Internet Use by Education, Percentage of Individuals



Source: File, Thom. 2013. Computer and Internet use in the United States. Current Population Survey Reports, P20-568. U.S. Census Bureau, Washington, DC.

People come to mistaken conclusions about the social world for various reasons: It’s easy to make errors in logic, particularly when we are analyzing the world in which we ourselves are conscious—and self-interested—participants. We can call some of these errors *everyday errors*, because they occur so frequently in the nonscientific, unreflective conversations that we hear on a daily basis.

A nice example of such errors in reasoning comes from a letter to Ann Landers, a popular newspaper advice columnist in the late 1900s. See if you can spot the problems here: The letter was written by a woman who had just moved, with her two pet cats, from an apartment in the city to a house in the country. In the city, she had not let the cats go outside, but she felt guilty about keeping them locked up. Upon arrival at the country house, she let the cats out—but they tiptoed cautiously to the door, looked outside, then went right back into the living room and lay down.

The woman concluded that people shouldn't feel guilty about keeping their cats indoors because even when they have the chance, cats don't really want to play outside.

Did you spot this person's errors in reasoning?

- *Overgeneralization*—She observed only two cats, both of which were previously confined indoors. Maybe they aren't like most cats.
- *Selective or inaccurate observation*—She observed the cats at the outside door only once. But maybe if she let them out several times, they would become more comfortable with going out.
- *Illogical reasoning*—She assumed that other people feel guilty about keeping their cats indoors. But maybe they don't.
- *Resistance to change*—She was quick to conclude that she had no need to change her approach to the cats. But maybe she just didn't want to change her own routines and was eager to believe that she was managing her cats just fine already.

You don't have to be a scientist or use sophisticated research techniques to avoid these four errors in reasoning. If you recognize them and make a conscious effort to avoid them, you can improve your own reasoning. In the process, you also will be taking the advice of your parents (or minister, teacher, or other adviser) not to stereotype people, to avoid jumping to conclusions, and to look at the big picture. These are the same kinds of mistakes that the methods of social science are designed to help us avoid.

Let's look at each kind of error in turn.

Overgeneralization

Overgeneralization occurs when we unjustifiably conclude that what is true for some cases is true for all cases. We are always drawing conclusions about people and social processes from our own interactions with them, but sometimes we forget that our experiences are limited. The social (and natural) world is, after all, a complex place. Maybe someone made a wisecrack about the ugly shoes you're wearing today, but that doesn't mean that "everyone is talking about you." Or there may have been two drunk-driving accidents following fraternity parties this year, but by itself, this doesn't mean that all fraternity brothers are drunk drivers. Or maybe you had a boring teacher in your high school chemistry class, but that doesn't mean all chemistry teachers are boring. We can interact with only a small fraction of the individuals who inhabit the social world, especially in a limited span of time; rarely are they completely typical people. One heavy Internet user found that his online friendships were "much deeper and have better quality" than his other friendships (Parks & Floyd 1996). Would his experiences generalize to yours? To those of others?

Overgeneralization: Occurs when we unjustifiably conclude that what is true for some cases is true for all cases.

Selective or Inaccurate Observation

We also have to avoid**selective** or **inaccurate observation**—choosing to look only at things that are in line with our preferences or beliefs. When we dislike individuals or institutions, it is all too easy to notice their every failing. For example, if we are convinced that heavy Internet users are antisocial, we can find many confirming instances. But what about elderly people who serve as Internet pen pals for grade school children or therapists who deliver online counseling? If we acknowledge only the instances that confirm our predispositions, we are victims of our own selective observation. Exhibit 1.2 depicts the difference between selective observation and overgeneralization.

Our observations can also simply be inaccurate. When you were in high school, maybe your mother complained that you were "always staying out late with your friends." Perhaps that was inaccurate; you only stayed out late occasionally. And when you complained that she "yelled" at you, even though her voice never actually increased in volume, that, too, was an inaccurate observation. In social science, we try to be more precise than that.

Such errors often occur in casual conversation and in everyday observation of the world around us. What we think we have seen is not necessarily what we really have seen (or heard, smelled, felt, or tasted). Even when our senses are functioning fully, our minds have to interpret what we have sensed (Humphrey 1992). The optical illusion in Exhibit 1.3, which can be viewed as either two faces or a vase, should help you realize that even simple visual perception requires interpretation.

Selective (inaccurate) observation: Choosing to look only at things that are in line with our preferences or beliefs.

Illogical Reasoning

When we prematurely jump to conclusions or argue on the basis of invalid assumptions, we are using **illogical reasoning**. For example, we might think that people who don't have many social ties just aren't friendly, even if we know they have just moved into a community and started a new job. Obviously, that's not logical. Conversely, an unquestioned assumption that everyone seeks social ties or benefits from them overlooks some important considerations, such as the impact of childhood difficulties on social trust and the exclusionary character of many tightly knit social groups. Logic that seems impeccable to one person can seem twisted to another—but people having different assumptions, rather than just failing to "think straight," usually causes the problem.

Exhibit 1.2 The Difference Between Overgeneralization and Selective Observation

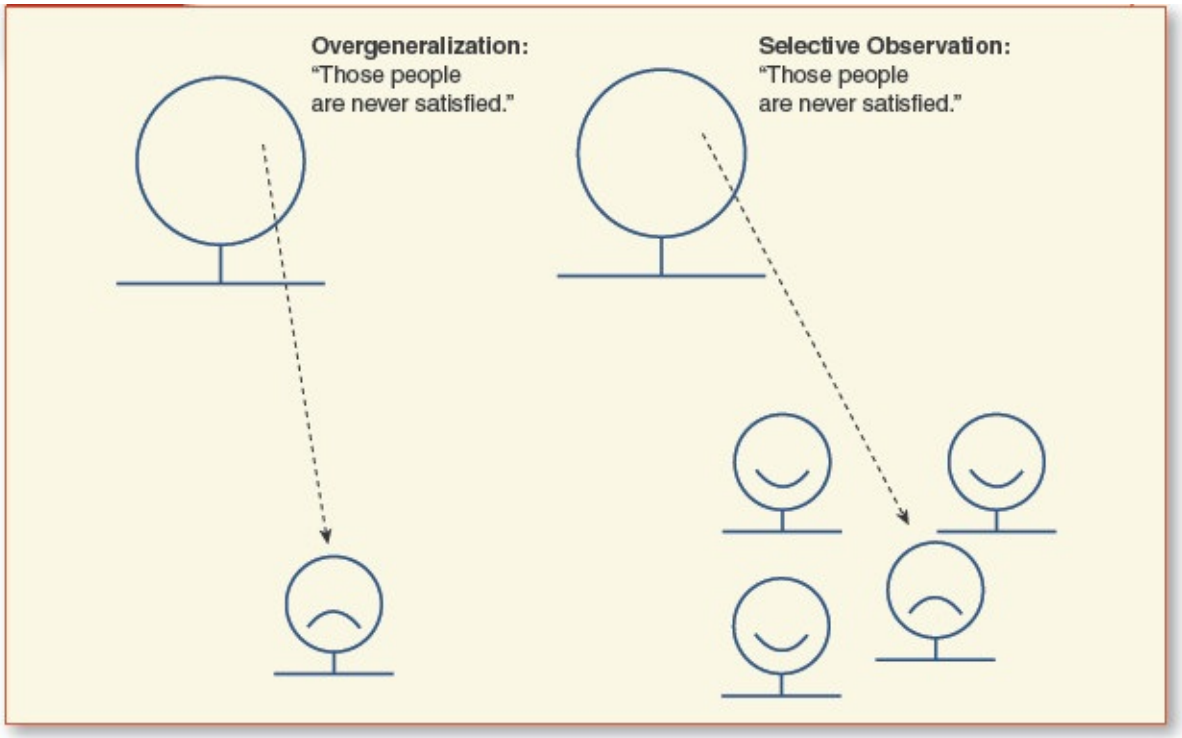
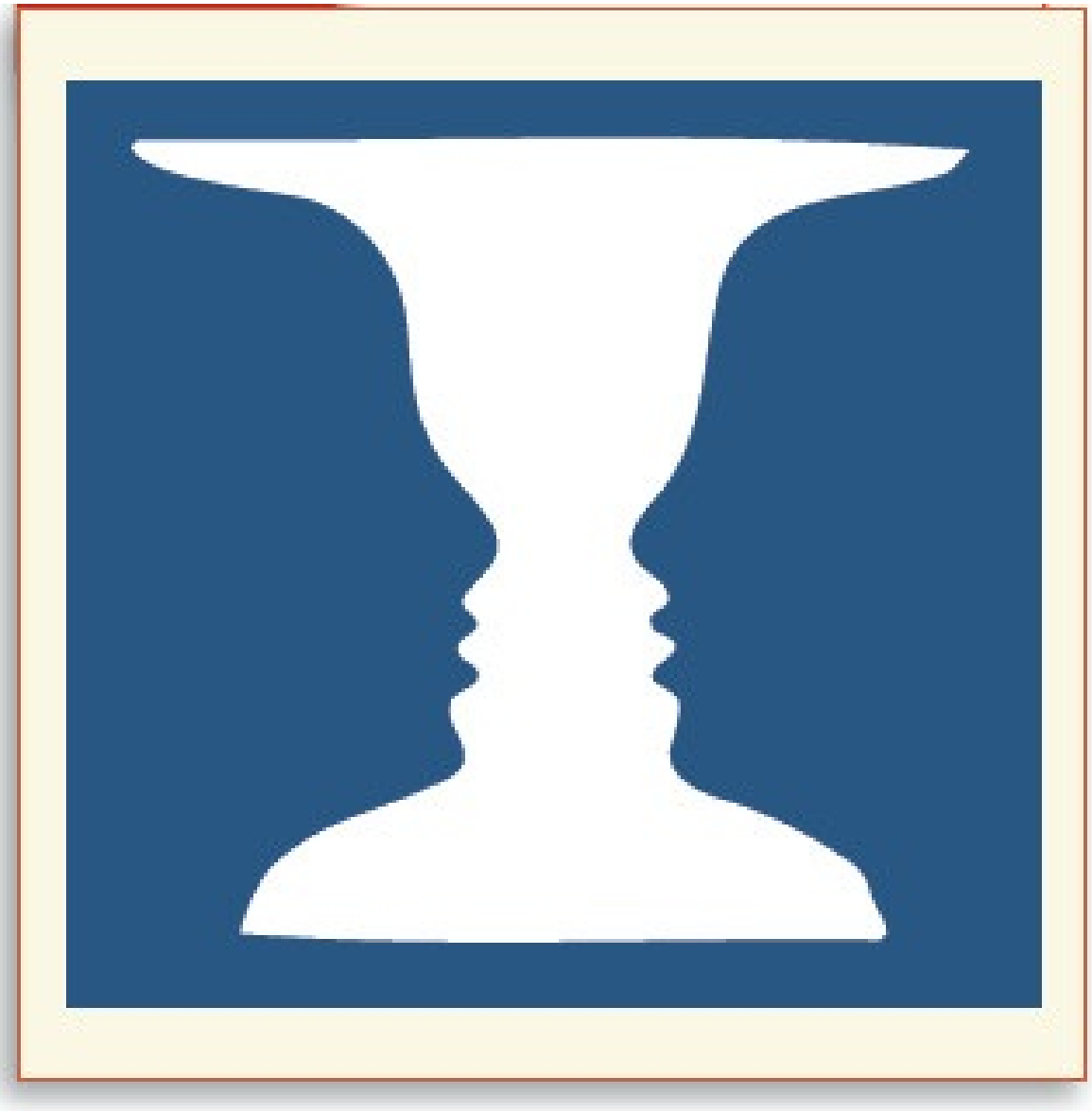


Exhibit 1.3 An Optical Illusion



Illogical reasoning: The premature jumping to conclusions or arguing on the basis of invalid assumptions.

Resistance to Change

Resistance to change, the reluctance to change our ideas in light of new information, is a common problem. After all, we know how tempting it is to make statements that conform to our own needs rather than to the observable facts (“I can’t live on that salary!”). It can also be difficult to admit that we were wrong once we have staked out a position on an issue (“I don’t want to discuss this anymore.”). Excessive devotion to tradition can stifle adaptation to changing circumstances (“This is how we’ve always done it, that’s why.”). People often accept the recommendations of those in positions of authority without question (“Only the president has all the facts.”). In all of these ways, we often close our eyes to what’s actually happening in the world.

Resistance to change: The reluctance to change our ideas in light of new information.

Research That Matters

How does wireless access to the Internet affect social life? Do people become less engaged with those around them? Will local community ties suffer? Since the development of the Internet in the 1980s, social scientists have been concerned with the impact of Internet connections on social interaction. Professor Keith Hampton at the University of Pennsylvania and Neeti Gupta at Microsoft studied wireless Internet users in four coffee shops in Boston and Seattle. They observed at each café for 30 hours, recording notes on the mobile device users' gender and approximate age as well as on their interaction with customers and staff. Hampton and Gupta concluded that there were two types of Internet users in the coffee shops: some were "true mobiles" who used the coffee shop as a place to work, for temporary or specific periods, and were largely disengaged from others around them. Others—"placemakers"—were primarily in the coffee shops to "hang out" and were very available for unplanned discussions with others about shared interests.

Source: Hampton, Keith N., and Neeti Gupta. 2008. Community and social interaction in the wireless city: Wi-Fi use in public and semi-public spaces. *New Media & Society* 10(6): 831–850.

Can Social Scientists See the Social World More Clearly?

Can social science do any better? Can we see the social world more clearly if we use the methods of social science? **Science** relies on logical and systematic methods to answer questions, and it does so in a way that allows others to inspect and evaluate its methods. So social scientists develop, refine, apply, and report their understanding of the social world more systematically, or "scientifically," than the general public does.

- **Social science** research methods reduce the likelihood of overgeneralization by using systematic procedures for selecting individuals or groups to study so that the study subjects are representative of the individuals or groups to which we want to generalize.
- To avoid illogical reasoning, social researchers use explicit criteria for identifying causes and for determining whether these criteria are met in a particular instance.
- Social science methods can reduce the risk of selective or inaccurate observation by requiring that we measure and sample phenomena systematically.
- Scientific methods lessen the tendency to answer questions about the social world from ego-based commitments, excessive devotion to tradition, or unquestioning respect for authority. Social scientists insist, Show us the evidence!

Social Research in Practice

Although all social science research seeks to minimize errors in reasoning, different projects may have different goals. The four most important goals of social research are (1) description, (2) exploration, (3) explanation, and (4) evaluation. Let's look at examples of each.

Journal Link

Read about our changing perspectives of community.

Description: How Often Do Americans "Neighbor"?

During the last quarter of the 20th century, the annual (biennial since 1996) General Social Survey (GSS) investigated a wide range of characteristics, attitudes, and behaviors. Each year, more than 1,000 adults in the United States completed GSS phone interviews; many questions repeated from year to year so that trends could be identified. Robert Putnam often used GSS data in his famous *Bowling Alone* investigation of social ties in America.

Researcher Interview Link

Read about social research in practice.

Survey responses indicated that "neighboring" declined throughout this period. As indicated in Exhibit 1.4 (Putnam 2000: 106), the percentage of GSS respondents who reported spending "a social evening with someone who lives in your neighborhood . . . about once a month or more often" was 60% for married people in 1975 and about 65% for singles. By 1998, the comparable percentages were 45% for married people and 50% for singles. This is **descriptive research** because the findings simply describe differences or variations in social phenomena.

Audio Link

Listen to other research by Putnam.

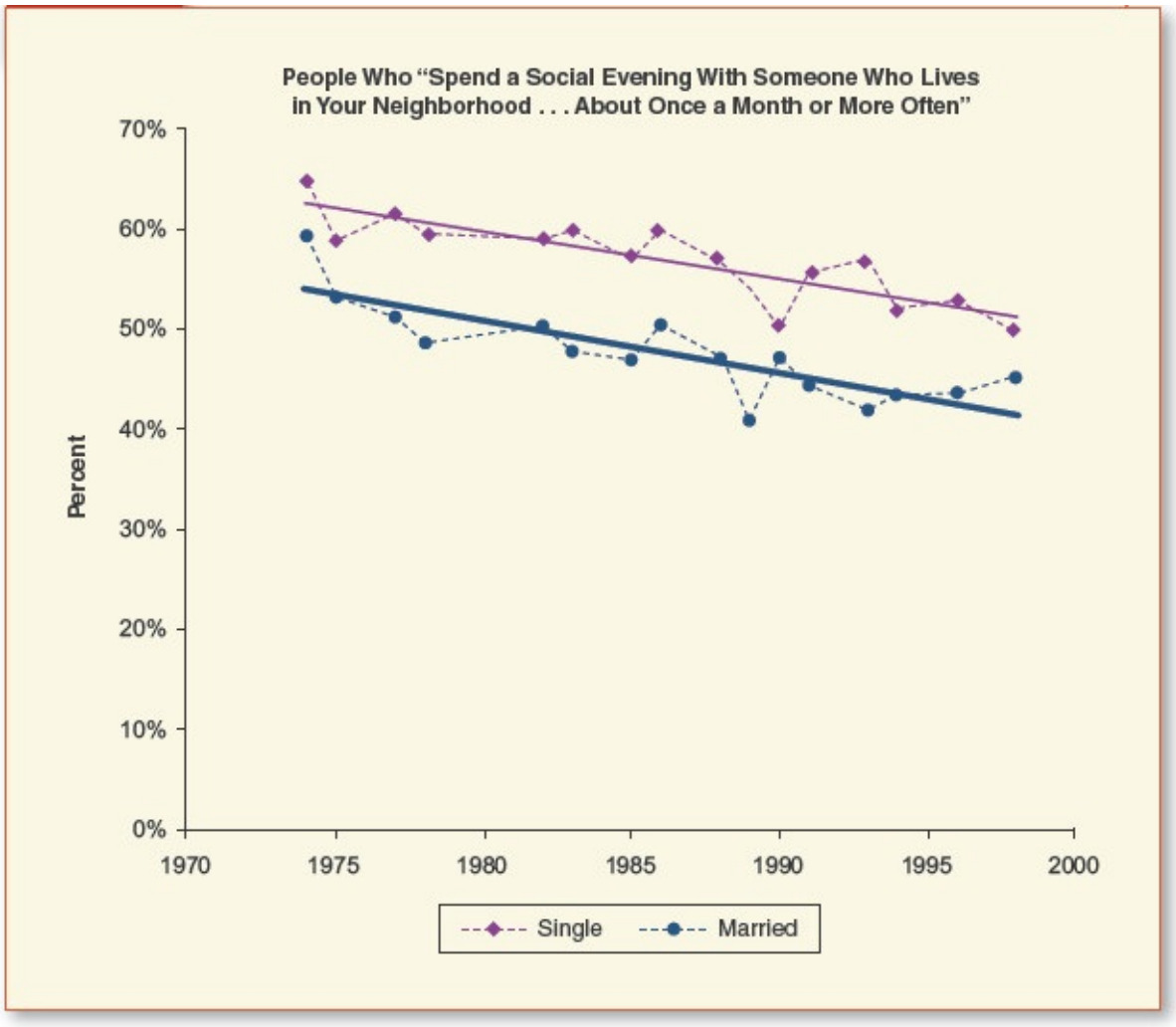
Science: A set of logical, systematic, documented methods for investigating nature and natural processes; the knowledge produced by these investigations.

Social science: The use of scientific methods to investigate individuals, societies, and social processes; the knowledge produced by these investigations.

Exploration: How Do Athletic Teams Build Player Loyalty?

Organizations such as combat units, surgical teams, and athletic teams must develop intense organizational loyalty among participants if organizations are to maximize their performance. How do they do it? This question motivated Patricia and Peter Adler (2000) to study college athletics. They wanted to explore this topic without preconceptions or fixed hypotheses. So Peter Adler joined his college basketball team as a "team sociologist," while Patti participated in some team activities as his wife and as a professor at the school. They recorded observations and comments at the end of each day for a period of 5 years. They also interviewed at length the coaches and all 38 basketball team members during that period.

Exhibit 1.4 The Decline of Neighboring 1974–1998



Source: Reprinted with permission of Simon & Schuster, Inc. from Bowling Alone by Robert D. Putnam. Copyright © 2000 Robert. D. Putnam.

Careful and systematic review of their notes led Adler and Adler (2000) to conclude that intense organizational loyalty emerged from five processes: (1) domination, (2) identification, (3) commitment, (4) integration, and (5) goal alignment. We won't review each of these processes here, but the following quote indicates how they found the process of integration into a cohesive group to work:

> By the time the three months were over [the summer before they started classes] I felt like I was there a year already. I felt so connected to the guys.

> You've played with them, it's been 130 degrees in the gym, you've elbowed each other, knocked each other around. Now you've felt a relationship, it's a team, a brotherhood type of thing. Everybody's got to eat the same rotten food, go through the same thing, and all you have is each other. So you've got a shared bond, a camaraderie. It's a whole houseful of brothers. And that's home to everybody in the dorm, not your parents' house. (p. 43)

Participating in and observing the team over this long period enabled Adler and Adler (2000) to identify and to distinguish particular aspects of such loyalty-building processes, such as identifying three modes of integration into the group: (1) unification in opposition to others, (2) development of group solidarity, and (3) sponsorship by older players. Adler and Adler also identified negative consequences of failures in group loyalty, such as the emergence of an atmosphere of jealousy and mistrust, and the disruption of group cohesion, as when one team member focused only on maximizing his own scoring statistics.

Researcher Interview Link

Read more about exploratory research.

In this project, Adler and Adler did more than simply describe what people did—they tried to explore the different elements of organizational loyalty and the processes by which loyalty was built. **Exploratory research** seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them. You might say the goal is to learn "what's going on here?"

Descriptive research: Research in which social phenomena are defined and described.

Exploratory research: Seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them.

Research in the News

Why Doesn't the Internet Reach Everyone?

As job applications, health care, movie viewing, and educational programs move online, the importance of access to high-speed Internet is increasing, but more than 100 million people are being left behind. The Department of Commerce reported that only 40% of households making $25,000 or less have Internet access at home. The problem of access is highly associated with the high cost of Internet contracts, the limited competition among Internet providers, and the lack of regulatory policies.

1. What else would you like to research to better understand the problem of those "left behind"?
2. How could you test the impact of lowering costs or of changing regulatory policies?

News Source: Crawford, Susan P. 2011. The new digital divide. *New York Times*, December 4: A1.

Explanation: Does Social Context Influence Adolescent Outcomes?

Often, social scientists want to explain social phenomena, usually by identifying causes and effects. Bruce Rankin at Koç University in Turkey and James Quane at Harvard University (Rankin & Quane 2002) analyzed data collected in a large survey of African American mothers and their adolescent children to test the effect of social context on adolescent outcomes. The source of data was a study funded by the MacArthur Foundation, Youth Achievement and the Structure of Inner City Communities, in which face-to-face interviews were conducted with more than 636 youth living in 62 poor and mixed-income urban Chicago neighborhoods.

Explanatory research like this seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in another phenomenon. Rankin and Quane (2002) were most concerned with determining the relative importance of three different aspects of social context—neighborhoods, families, and peers—on adolescent outcomes (both positive and negative). To make this determination, they had to conduct their analysis in a way that allowed them to separate the effects of neighborhood

characteristics, such as residential stability and economic disadvantage, from parental involvement in child rearing and other family features, as well as from peer influence. They found that neighborhood characteristics affect youth outcomes primarily by influencing the extent of parental monitoring and the quality of peer groups.

Journal Link

Read about how social interventions can affect well-being.

> **Explanatory research:** Seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in another phenomenon.

Evaluation: Does More Social Capital Result in More Community Participation?

The “It’s Our Neighbourhood’s Turn” project (Onze Buurt aan Zet, or OBAZ) in the city of Enschede, the Netherlands, was one of a seri es of projects initiated by the Dutch Interior and Kingdom Relations ministry to increase the quality of life and safety of individuals in the most deprived neighborhoods in the Netherlands. In the fall of 2001, residents in three of the city’s poorest neighborhoods were informed that their communities had received funds to use for community improvement and that residents had to be actively involved in formulating and implementing the improvement plans (Lelieveldt 2003: 1). Political scientist Herman Lelieveldt (2004: 537) at the University of Twente, the Netherlands, and others then surveyed community residents to learn about their social relations and their level of local political participation; a second survey was conducted 1 year after the project began.

Lelieveldt wanted to evaluate the impact of the OBAZ project—to see whether the “livability and safety of the neighborhood” could be improved by taking steps like those Putnam (2000: 408) recommended to increase “social capital,” meaning that citizens would spend more time connecting with their neighbors.

Encyclopedia Link

Read about evaluation research.

It turned out that residents who had higher levels of social capital participated more in community political processes. However, not every form of social capital made much of a difference. Neighborliness—the extent to which citizens are engaged in networks with their neighbors—was an important predictor of political participation, as was a feeling of obligation to participate. By contrast, a sense of trust in others (something that Putnam emphasizes) was not consistently important (Lelieveldt 2004: 535, 547–548): Those who got more involved in the OBAZ political process tended to distrust their neighbors. When researchers focus their attention on social programs such as the OBAZ project, they are conducting **evaluation research**—research that describes or identifies the impact of social policies and programs.

Interactive Exercises Link

Types of Research

Certainly many research studies have more than one such goal—all studies include some description, for instance. But clarifying your primary goal can often help when deciding how to do your research.

Evaluation research: Research that describes or identifies the impact of social policies and programs.

Careers and Research

Jessica LeBlanc, Research Assistant

Jessica LeBlanc majored in sociology at the University of New Hampshire, but she didn't really know what kind of career it would lead to. Then she took an undergraduate statistics course and found she really enjoyed it. She took additional methods courses—survey research and an individual research project course—and really liked those also.

By the time she graduated, LeBlanc knew she wanted a job in social research. She looked online for research positions in marketing, health care, and other areas. She noticed an opening at the Center for Survey Research (CSR) at the University of Massachusetts in Boston and thought their work sounded fascinating. The job description said "MA preferred," but within a week she had an interview and then was hired. LeBlanc liked CSR because it was academic, had a wide range of projects, and had many that were focused on her primary interests in health.

As a Research Assistant II, LeBlanc designed survey questions, transcribed focus group audiotapes, programmed web surveys, and managed incoming data. She also conducted focus groups and interviews and programmed computer-assisted telephone surveys.

The knowledge that LeBlanc gained in her methods courses about research designs, statistics, question construction, and survey procedures prepared her well for her position at CSR. She has found that it's important to understand validity and reliability and the basics of statistical software. Her advice to aspiring researchers: Pay attention in your first methods class!

LeBlanc has also benefited from on-the-job training. In her first year, she learned the ins and outs of the center and social research, she completed an online course in human subjects protections, and she learned how to conduct cognitive interviews and moderate focus groups. She's also learned how to use Microsoft Access and Excel and how to program surveys delivered through computers. Overall, LeBlanc enjoys the nitty-gritty and hands-on, day-to-day management task.

How Well Have We Done Our Research?

Social scientists want validity in their research findings—they want to find the truth. The goal of social science is not to reach conclusions that other people will like or that suit our personal preferences. We shouldn't start our research determined to "prove" that our college's writing program is successful, or that women are portrayed unfairly in advertisements, or that the last presidential election was rigged, or that homeless people are badly treated. We may learn that all of these are true, or aren't, but our goal as social scientists should be to learn the truth, even if it's sometimes disagreeable to us. The goal is to figure out how and why some part of the social world operates as it does and to reach valid conclusions. We reach the goal of **validity** when our statements or conclusions about empirical reality are correct. In *Making Sense of the Social World: Methods of Investigation*, we will be concerned with three kinds of validity: (1) measurement validity, (2) generalizability, and (3) causal validity (also known as internal validity). We will learn that invalid measures, invalid generalizations, or invalid causal inferences result in invalid conclusions.

Measurement Validity

Measurement validity is our first concern because without having measured what we think we've measured, we don't even know what we're talking about. So when Putnam (2000: 291) introduces a measure of "social capital" that has such components as number of club meetings attended and number of times worked on a community project, we have to stop and consider the validity of this measure. Measurement validity is the focus of Chapter 4.

Problems with measurement validity can occur for many reasons. In studies of Internet forums, for instance, researchers have found that some participants use fictitious identities, even pretending to be a different gender (men posing as women, for instance) (Donath 1999). Therefore, it's difficult to measure gender in these forums, and researchers could not rely on gender as disclosed in the forums when identifying differences in usage patterns between men and women. Similarly, if you ask people, "Are you an alcoholic?" they probably won't say yes, even if they are; the question elicits less valid information than would be forthcoming by asking them how many drinks they consume, on average, each day. Some college students may be hesitant to admit they binge-watch *Breaking Bad* on television 6 hours a day, so researchers use electronic monitoring devices on TV sets to measure what programs people watch and how often.

Encyclopedia Link

Read an overview of the importance of generalizability.

Validity: The state that exists when statements or conclusions about empirical reality are correct.

Measurement validity: Exists when an indicator measures what we think it measures.

Generalizability

The **generalizability** of a study is the extent to which it can inform us about persons, places, or events that were not directly studied. For instance, if we ask our favorite students how much they enjoyed our Research Methods course, can we assume that other students (perhaps not as favored) would give the same answers? Maybe they would—but probably not. Achieving generalizability through correct sampling is the focus of Chapter 5.

Generalizability is always an important consideration when you review social science research. Even the huge, international National Geographic Society (2000) survey of Internet users had some limitations in generalizability. Only certain people were included in the sample: people who were connected to the Internet, who had heard about the survey, and who actually chose to participate. This meant that many more respondents came from wealthier countries, which had higher rates of computer and Internet use, than from poorer countries. However, the inclusion of individuals from 178 countries and territories does allow some interesting comparisons among countries.

There are two kinds of generalizability: sample and cross-population.

Sample generalizability is a key concern in survey research. Political polls, such as the Gallup Poll or Zogby International, may study a sample of 1,400 likely voters, for example, and then generalize the findings to the entire American population of 120 million likely voters. No one would be interested in the results of political polls if they represented only the tiny sample that actually was surveyed rather than the entire population.

Cross-population generalizability occurs to the extent that the results of a study hold true for multiple populations; these populations may not all have been sampled, or they may be represented as subgroups within the sample studied (see Exhibit 1.5). We can only wonder about the cross-population generalizability of Putnam's findings about social ties in the United States. Has the same decline occurred in Mexico, Argentina, Britain, or Thailand?

Generalizability: Exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify; it is the extent to which a study can inform us about persons, places, or events that were not directly studied.

Sample generalizability: Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

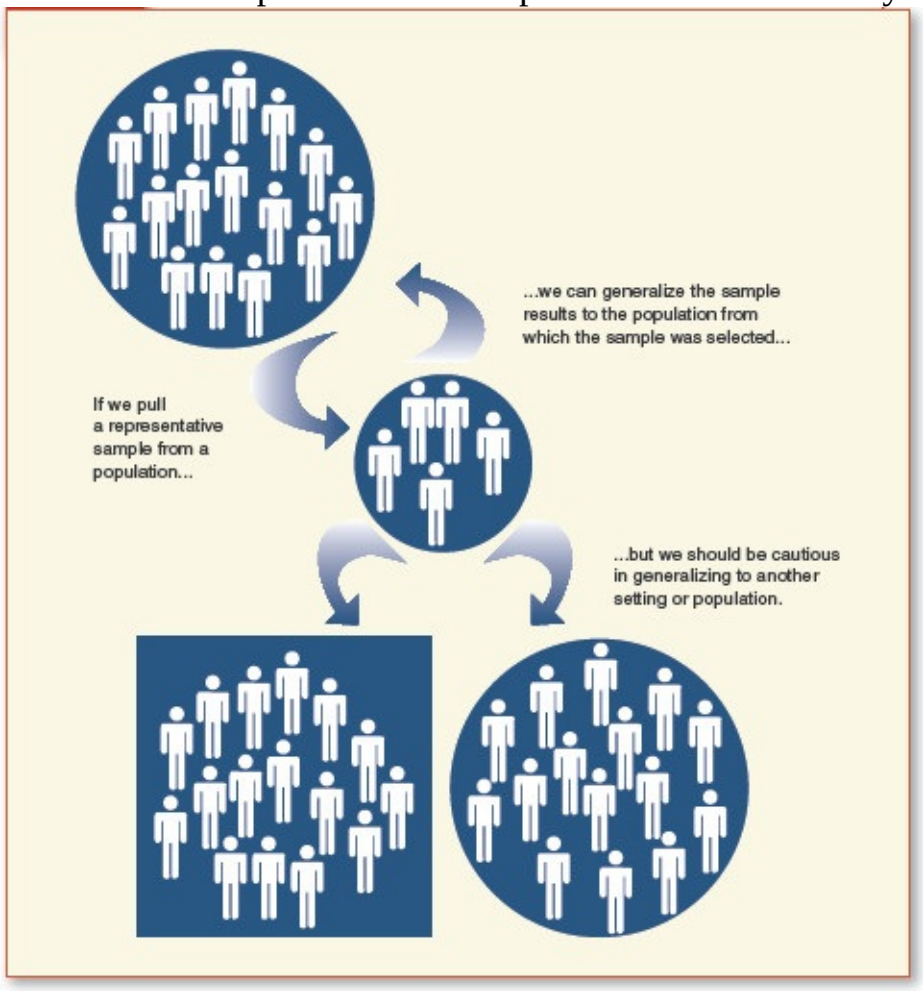
Cross-population generalizability (external validity): Exists when findings about one group, population, or setting hold true for other groups, populations, or settings.

Causal validity (internal validity): Exists when a conclusion that A leads to, or results in, B is correct.

Causal Validity

Causal validity, also known as **internal validity**, refers to the truthfulness of an assertion that A causes B. It is the focus of Chapter 6.

Exhibit 1.5 Sample and Cross-Population Generalizability


...we can generalize the sample results to the population from which the sample was selected...
If we pull a representative sample from a population...
...but we should be cautious in generalizing to another setting or population.

Most research seeks to determine what causes what, so social scientists frequently must be concerned with causal validity. For example, Gary Cohen and Barbara Kerr (1998) asked whether computer-mediated counseling could be as effective as face-to-face counseling for mental health problems—that is, whether one type of counseling leads to better results than the other. Cohen and Kerr could have compared people who had voluntarily experienced one of these types of treatment, but it's quite likely that individuals who sought out a live person for counseling would differ, in important ways, from those who sought

computer-mediated counseling. Younger people tend to use computers more; so do more educated people. Or maybe less sociable people would be more drawn to computer-mediated counseling. Normally, it would be hard to tell if different results from the two therapies were caused by the therapies themselves or by different kinds of people going to each.

So Cohen and Kerr (1998) designed an experiment in which students seeking counseling were assigned randomly (by a procedure somewhat like flipping a coin) to either computer-mediated or face-to-face counseling. In effect, people going to one kind of counseling were just like people going to the other; as it happens, their anxiety scores afterward were roughly the same. There seemed to be no difference (Exhibit 1.6). By using the random assignment procedure, Cohen and Kerr strengthened the causal validity of this conclusion.

Journal Link

Read about data based on a representative sample.

Conversely, even in properly randomized experiments, causal findings can be mistaken because of some factor that was not recognized during planning for the study. If the computer-mediated counseling sessions were conducted in a modern building with all the latest amenities, but face-to-face counseling was delivered in a run-down building, this difference might have led to different outcomes for reasons quite apart from the type of counseling. Also, Cohen and Kerr didn't have a group that received no counseling. Maybe just a little quiet time or getting older would provide the same benefits as therapy.

So establishing causal validity can be quite difficult. In subsequent chapters, you will learn in more detail how experimental designs and statistics can help us evaluate causal propositions, but the solutions are neither easy nor perfect. We always have to consider critically the validity of causal statements that we hear or read.

Video Link

Watch a panel discussion on the impacts and use of social research.

Conclusion

This first chapter should have given you an idea of what to expect in the rest of the book. Social science provides us with a variety of methods for avoiding everyday errors in reasoning and for coming to valid conclusions about the social world. We will explore different kinds of research, using different techniques, in the chapters to come, always asking, is this answer likely to be correct? The techniques are fairly simple, but they are powerful nonetheless if properly executed. You will also learn some interesting facts about social life. We have already seen, for instance, some evidence that

Exhibit 1.6 Partial Evidence of Causality

Precounseling Anxiety Score	Type of Counseling	Postcounseling Anxiety Score
35	Computer-mediated	28
35	Face-to-face	29

Precounseling anxiety score: 35 — Computer-mediated counseling → Postcounseling anxiety score: 28

Precounseling anxiety score: 35 — Face-to-face counseling → Postcounseling anxiety score: 29

- The Internet and social media may have surprising effects on our relationships with others.
- Organizational processes that build loyalty, as happens on athletic teams, can strengthen social ties.
- Neighborhoods in which social ties are weaker may result in less effective

forms of parenting, but both parenting and peer group quality have stronger effects than neighborhood social ties on adolescent outcomes.
- Government programs to increase social capital in neighborhoods can increase local political participation.
- Students may benefit as much from computer-mediated counseling as from face-to-face counseling.

Remember, you must ask a direct question of each research project you examine: How valid are its conclusions? The theme of validity ties the chapters in this book together. Each technique will be evaluated for its ability to help us with measurement validity, generalizability, and causal validity.

To illustrate the process of doing research, in Chapter 2, we describe studies of domestic violence, community disaster, student experience of college, and other topics. We review the types of research questions that social scientists ask, the role of theory, the major steps in the research process, and other sources of information that may be used in social research. In Chapter 3, we set out the general principles of ethical research that social scientists try to follow. As well, examples of ethical challenges to good research will be presented in many of the chapters that follow.

Then in Chapters 4, 5, and 6, we return to the subject of validity—the three kinds of validity and the specific techniques used to maximize the validity of our measures, our generalizations from a sample, and our causal assertions. Chapter 6 also introduces experimental studies, one of the best methods for establishing causal relationships.

Other methods of data collection and analysis are introduced in Chapters 7, 8, 9, and 10. Survey research is the most common method of data collection in sociology, and in Chapter 7, we devote attention to the different types of surveys. Chapter 8 is not a substitute for an entire course in statistics, but it gives you a good idea of how to use statistics honestly in reporting the results of your own studies using quantitative methods, in interpreting the results of research reported by others, and in analyzing secondary data sources. Chapter 9 shows how qualitative methods such as participant observation, intensive interviewing, and focus groups can uncover aspects of the social world that we are likely to miss in experiments and surveys, and Chapter 10, on qualitative data analysis, illustrates several approaches that researchers can take to the analysis of the data they collect in qualitative projects.

Chapter 11 introduces a range of unobtrusive measures that aren't experienced by the people being studied; these include historical and comparative methods, content analysis, and a variety of creative techniques. Chapter 12 explains the role of evaluation research in investigating social programs and how to design evaluation research studies. Finally, Chapter 13 focuses on how to review prior research, how to propose new research, and how to report original research. We give special attention to how to formulate research proposals and how to critique, or evaluate, reports of research that you encounter.

Throughout these chapters, we will try to make the ideas interesting and useful to you, both as a consumer of research (as reported in newspapers, for instance) and as a potential producer (if, say, you do a survey in your college, neighborhood, or business). Each chapter ends with several helpful learning tools. Lists of key terms and chapter highlights will help you review, and exercises will help you apply your knowledge. Social research isn't rocket science, but it does take some clear thinking, and these exercises should give you a chance to practice.

Here is a closing thought: Vince Lombardi, legendary coach of the Green Bay Packers of the National Football League during the 1960s, used to say that championship football was basically a matter of "four yards and a cloud of dust." Nothing too fancy, no razzle-dazzle plays, no phenomenally talented players doing it all alone—just solid, hard-working, straight-ahead fundamentals. This may sound strange, but excellent social research can be done—can "win games"—in the same way. We'll show you how to design and conduct surveys that get the right answers, interviews that discover people's true feelings, and experiments that pinpoint what causes what. And we'll show you how to avoid getting taken in by every "Studies Show . . . We're Committing More Crimes!" article you read on the Internet. It takes a little effort initially, but we think you will find it worthwhile—even enjoyable.

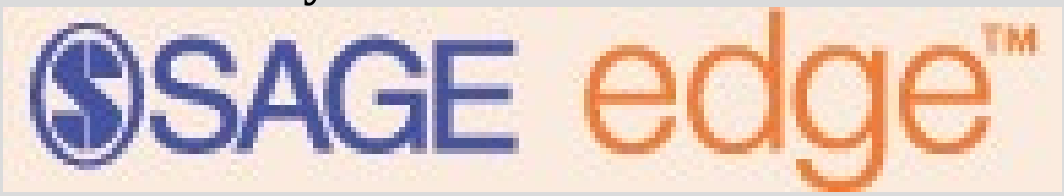
Key Terms

Causal validity (internal validity) 12
Cross-population generalizability (external validity) 12
Descriptive research 8
Evaluation research 10
Explanatory research 10
Exploratory research 9
Generalizability 12
Illogical reasoning 5
Measurement validity 12
Overgeneralization 5
Resistance to change 6
Sample generalizability 12
Science 7
Selective (or inaccurate) observation 5
Social science 7
Validity 11

Highlights

- Four common errors in everyday reasoning are overgeneralization, selective or inaccurate observation, illogical reasoning, and resistance to change. These errors result from the complexity of the social world, subjective processes that affect the reasoning of researchers and those they study, researchers' self-interestedness, and unquestioning acceptance of tradition or of those in positions of authority.
- Social science is the use of logical, systematic, documented methods to investigate individuals, societies, and social processes, as well as the knowledge these investigations produce.
- Social research can be descriptive, exploratory, explanatory, or evaluative—or some combination of these.
- Valid knowledge is the central concern of scientific research. The three components of validity are measurement validity, generalizability (both from the sample to the population from which it was selected and from the sample to other populations), and causal (internal) validity.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Select a social issue that interests you, such as Internet use or crime. List at least four of your beliefs about this phenomenon. Try to identify the sources of each of these beliefs.
2. Does the academic motivation to do the best possible job of understanding how the social world works conflict with policy or personal motivations? How could personal experiences with social isolation or with Internet use shape research motivations? In what ways might the goal of influencing policy about social relations shape how a researcher approaches this issue?
3. Pick a contemporary social issue of interest to you. List descriptive, exploratory, explanatory, and evaluative questions that you could investigate about this issue.
4. Review each of the three sets of research alternatives. Which alternatives are most appealing to you? Which combination of alternatives makes the most sense to you (one possibility, for example, is quantitative research with a basic science orientation)? Discuss the possible bases of your research preferences relative to your academic interests, personal experiences, and policy orientations.

Finding Research

1. Read the abstracts (initial summaries) of each article in a recent issue of a major social science journal. (Ask your instructor for some good journal titles.) On the basis of the abstract only, classify each research project represented in the articles as primarily descriptive, exploratory, explanatory, or evaluative. Note any indications that the research focused on other types of research questions.
2. From the news, record statements of politicians or other leaders about some social phenomenon. Which statements do you think are likely to be in error? What evidence could the speakers provide to demonstrate the validity of these statements?
3. Check out Robert Putnam's website (www.bettertogether.org) and review survey findings about social ties in several cities. Prepare a 5- to 10-minute class presentation on what you have found about social ties and the ongoing research-based efforts to understand them.

Critiquing Research

1. Scan one of the publications about the Internet and society at the Berkman Center for Internet & Society website (http://cyber.law.harvard.edu/). Describe one of the projects discussed: its goals, methods, and major findings. What do the researchers conclude about the impact of the Internet on social life in the United States? Next, repeat this process with a report from the Pew Internet Project (www.pewinternet.org), or with the Digital Future report from the University of Southern California's Center for the Digital Future site (www.digitalcenter.org). What aspects of the methods, questions, or findings might explain differences in their conclusions? Do you think the researchers approached their studies with different perspectives at the outset? If so, what might these perspectives have been?
2. Research on social ties was publicized in a *Washington Post* article that also included comments by other sociologists (http://www.washingtonpost.com/wp-dyn/content/article/2006/06/22/AR2006062201763.html). Read the article, and continue the commentary. Do your own experiences suggest that there is a problem with social ties in your community? Does it seem, as Barry Wellman suggests in the *Washington Post* article, that a larger number of social ties can make up for the decline in intimate social ties that McPherson et al. (2006: 358) found?

Doing Research

1. What topic would you focus on if you could design a social research project without any concern for costs? What are your motives for studying this topic?
2. Develop four questions that you might investigate about the topic you just selected. Each question should reflect a different research goal: description, exploration, explanation, or evaluation. Be specific. Which question most interests you? Why?

Ethics Questions

Throughout the book, we will discuss the ethical challenges that arise in social research. At the end of each chapter, we ask you to consider some questions about ethical issues related to that chapter's focus. We introduce this critical topic formally in Chapter 3, but we begin here with some questions for you to ponder.

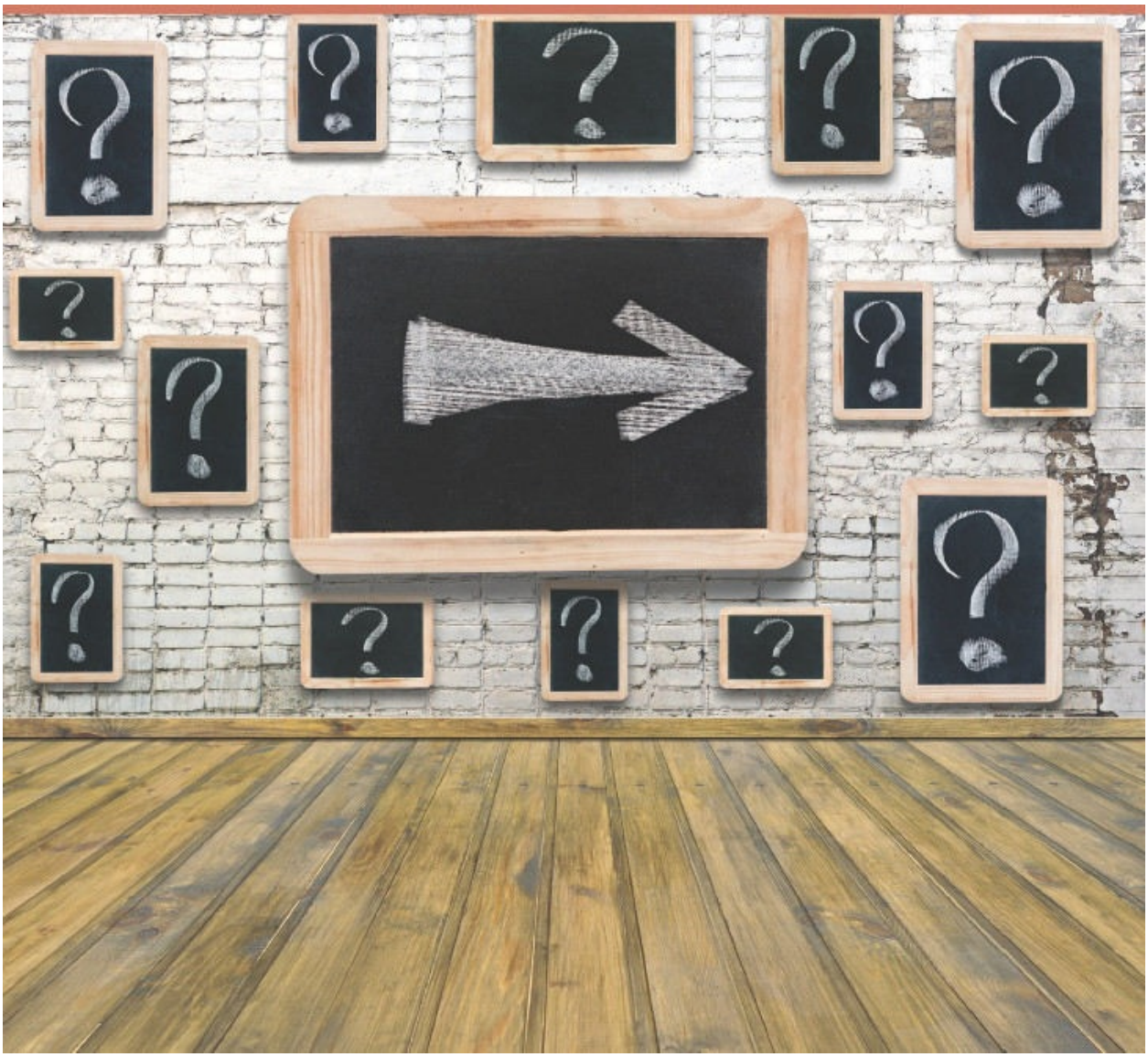
1. The chapter began with a brief description of research on social media and Internet use. What would you do if you were interviewing college students who spent lots of time online and found that some were very isolated and depressed or even suicidal, apparently as a result of the isolation? Do you believe that social researchers have an obligation to take action in a situation like this? What if you discovered a similar problem with a child? What guidelines would you suggest for researchers?
2. Would you encourage social researchers to announce their findings about problems such as social isolation in press conferences and to encourage relevant agencies to adopt policies encouraged to lessen social isolation? Should policies regarding attempts to garner publicity and shape policy depend on the strength of the research evidence? Do you think there is a fundamental conflict between academic and policy motivations? Do social researchers have an ethical obligation to recommend policies that their research suggests would help other people?

Video Interview Questions

Listen to the researcher interview for Chapter 1 at edge.sagepub.com/chamblissmssw5e.

1. What are the benefits to breaking down questions in text-based interview structure?
2. As Janet Salmons mentions, one can enhance his or her research by deciding carefully on the various kinds of technology to be used. What are some of considerations Salmons mentions in deciding whether to use text-based interviews or video conference calls?

Chapter 2 The Process and Problems of Social Research



©iStockphoto.com/inxti

Learning Objectives

1. Name the three characteristics of a good research question.
2. Define *theory*.
3. Contrast the process of research reflecting deductive reasoning with that reflecting inductive reasoning.
4. Understand why an explanation formulated after the fact is necessarily less certain than an explanation presented before the collection of data.
5. Diagram the research circle and explain the role of replication in relation to that circle.
6. Distinguish research designs using individuals and groups and explain their relation to the ecological and individualist fallacies.
7. Understand the differences between cross-sectional research designs and the three types of longitudinal research design.

In Chapter 1, we introduced the reasons *why* we do social research: to describe, explore, explain, and evaluate. Each type of social research can have tremendous impact. Alfred Kinsey's descriptive studies of the sex lives of Americans, conducted in the 1940s and 1950s, were at the time a shocking exposure of the wide variety of sexual practices that apparently staid, "normal" people engaged in behind closed doors—and the studies helped introduce the unprecedented sexual openness we see 70 years later (Kinsey, Pomeroy, & Martin 1948; Kinsey, Pomeroy, Martin, & Gebhard 1953). At around the same time, Gunnar Myrdal's exploratory book, *An American Dilemma* (1944/1964), forced our grandparents and great-grandparents to confront the tragedy of institutional racism. Myrdal's research was an important factor in the 1954 Supreme Court decision *Brown v. Board of Education of Topeka*, which ended school segregation in the United States. The explanatory *broken windows* theory of crime, which was developed during the 1980s by George Kelling and James Q. Wilson (1982), dramatically changed police practices in our major cities; its usefulness in reducing crime, and on feeding controversial "stop and frisk" programs, is hotly debated both in academic journals (Sampson and Raudenbusch 1999) and on the front pages of newspapers to this day (Goldstein 2014) And evaluative social research today actively influences advertising campaigns, federal housing programs, the organization of military units (from Army fire teams to Navy submarine crews), drug treatment programs, and corporate employee benefit plans.

Video Link

Watch some advice for new researchers.

We now introduce the *how* of social research. In this chapter, you will learn about the process of specifying a research question, developing an appropriate research strategy and design with which to investigate that question, and choosing appropriate units of analysis. By the chapter's end, you should be ready to formulate a question, to design a strategy for answering the question, and to begin to critique previous studies that addressed the question.

What Is the Question?

A **social research question** is a question about the social world that you seek to answer through the collection and analysis of firsthand, verifiable, empirical data. Questions like this may emerge from your own experience, from research by other investigators, from social theory, or from a *request for research* issued by a government agency that needs a study of a particular problem.

Some researchers of the health care system, for example, have had personal experiences as patients with serious diseases, as nurses or aides working in hospitals, or as family members touched directly and importantly by doctors and hospitals. These researchers may want to learn why our health care system failed or helped them. Feminist scholars study violence against women in hopes of finding solutions to this problem as part of a broader concern with improving women's lives. One colleague of ours, Veronica Tichenor, was fascinated by a prominent theory of family relations that argues that men do less housework than women do because men earn more money; Professor Tichenor did research on couples in which the woman made far more money than the man to test the theory. (She found, by the way, that the women still did more of the housework.) Some researchers working for large corporations or major polling firms conduct marketing studies simply to make money. So, a wide variety of motives can push a researcher to ask research questions.

A good research question doesn't just spring effortlessly from a researcher's mind. You have to refine and evaluate possible research questions to find one that is worthwhile. It's a good idea to develop a list of possible research questions as you think about a research area. At the appropriate time, you can narrow your list to the most interesting and feasible candidate questions.

What makes a research question "good"? Many social scientists evaluate their research questions in terms of three criteria: *feasibility* given the time and resources available, *social importance*, and *scientific relevance* (King, Keohane, & Verba 1994):

- Can you start and finish an investigation of your research question with available resources and in the time allotted? If so, your research question is feasible.
- Will an answer to your research question make a difference in the social

world, even if it only helps people understand a problem they consider important? If so, your research question is socially important.
- Does your research question help resolve some contradictory research findings or a puzzling issue in social theory? If so, your research question is scientifically relevant.

Here's a good example of a question that is feasible, socially important, and scientifically relevant: Does arresting accused spouse abusers on the spot prevent repeat incidents? Beginning in 1981, the Police Foundation and the Minneapolis Police Department began an experiment to find the answer. The Minneapolis experiment was first and foremost scientifically relevant: It built on a substantial body of contradictory theory regarding the impact of punishment on criminality (Sherman & Berk 1984). Deterrence theory predicted that arrest would deter individuals from repeat offenses, but labeling theory predicted that arrest would make repeat offenses more likely. The researchers found one prior experimental study of this issue, but it had been conducted with juveniles. Studies among adults had not yielded consistent findings. Clearly, the Minneapolis researchers had good reason for conducting a study.

As you consider research questions, you should begin the process of consulting and then reviewing the published literature. Your goal here and in subsequent stages of research should be to develop a research question and specific expectations that build on prior research and to use the experiences of prior researchers to chart the most productive directions and design the most appropriate methods. Appendix A describes how to search the literature, and Chapter 13 includes detailed advice for writing up the results of your search in a formal review of the relevant literature.

Encyclopedia Link

Read about how applied sociology helps make a difference.

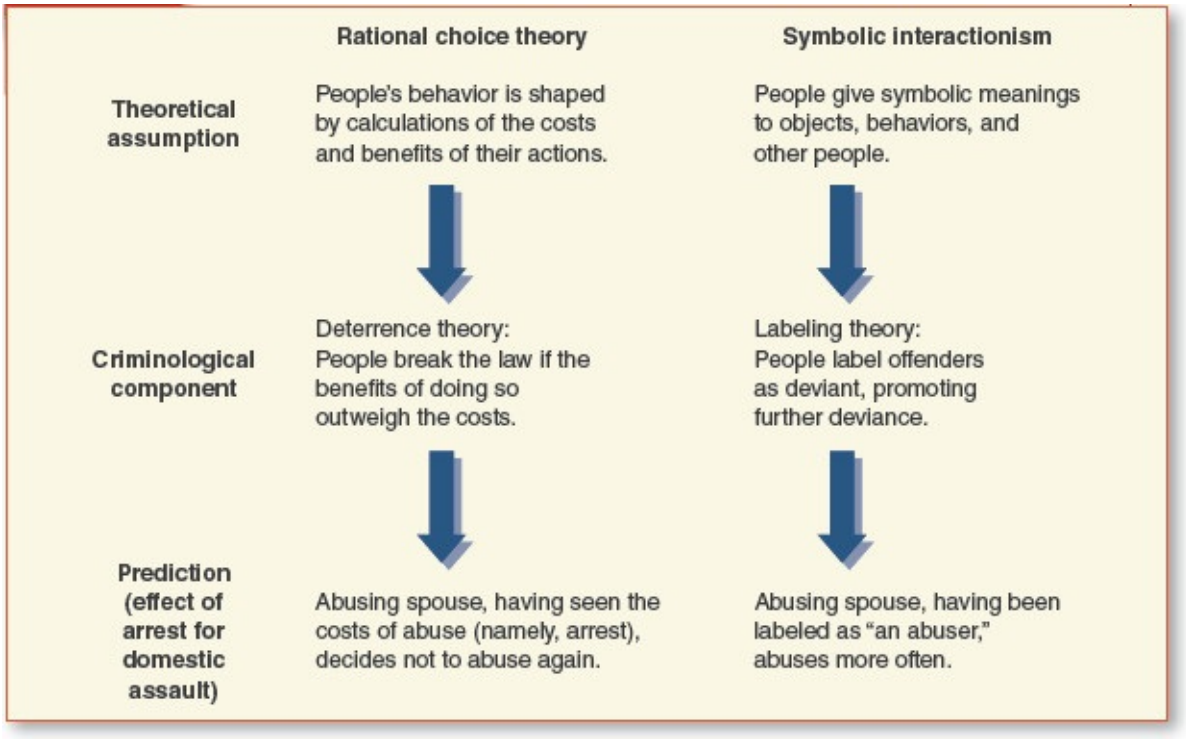
Social research question: A question about the social world that is answered through the collection and analysis of firsthand, verifiable, empirical data.

What Is the Theory?

Theories have a special place in social research because they help us make connections to general social processes and large bodies of research. Building and evaluating theory is, therefore, one of the most important objectives of social science. A social **theory** is a logically interrelated set of propositions about empirical reality (i.e., the social world as it actually exists). You may know, for instance, about conflict theory, which proposes that (1) people are basically self-interested, (2) power differences between people and groups reflect the different resources available to groups, (3) ideas (religion, political ideologies, etc.) reflect the power arrangements in a society, (4) violence is always a potential resource and the one that matters most, and so on (Collins 1975). These statements are related to each other, and the sum of conflict theory is a sizable collection of such statements (entire books are devoted to it). Dissonance theory in psychology, deterrence theory in criminology, and labeling theory in sociology are other examples of social theories.

Social theories suggest the areas on which we should focus and the propositions that we should consider testing. For example, Lawrence Sherman and Richard Berk's (1984) domestic violence research in the Minneapolis spouse abuse experiment was actually a test of predictions that they derived from two varying theories on the impact of punishment on crime (Exhibit 2.1).

Exhibit 2.1 Two Social Theories and Their Predictions About the Effect of Arrest on Domestic Assault


Rational choice theory
Symbolic interactionism
Theoretical assumption
People's behavior is shaped by calculations of the costs and benefits of their actions.
People give symbolic meanings to objects, behaviors, and other people.
Criminological component
Deterrence theory: People break the law if the benefits of doing so outweigh the costs.
Labeling theory: People label offenders as deviant, promoting further deviance.
Prediction (effect of arrest for domestic assault)
Abusing spouse, having seen the costs of abuse (namely, arrest), decides not to abuse again.
Abusing spouse, having been labeled as "an abuser," abuses more often.

Source: Data from Sherman, Lawrence W., and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. American Sociological Review 49: 267.

Deterrence theory expects punishment to deter crime in two ways. General deterrence occurs when people see that crime results in undesirable punishments —that "crime doesn't pay." The persons who are punished serve as examples of what awaits those who engage in proscribed acts. Specific deterrence occurs when persons who are punished decide not to commit another offense so they can avoid further punishment (Lempert & Sanders 1986: 86–87). Deterrence theory leads to the prediction that arresting spouse abusers will lessen their likelihood of reoffending.

Labeling theory distinguishes between primary deviance, the acts of individuals that lead to public sanction, and secondary deviance, the deviance that occurs in response to public sanction (Hagan 1994: 33). Arrest or some other public sanction for misdeeds labels the offender as deviant in the eyes of others. Once the offender is labeled, others will treat the offender as a deviant, and the offender is then more likely to act in a way that is consistent with the deviant label. Ironically, the act of punishment stimulates more of the very behavior that

it was intended to eliminate. This theory suggests that persons arrested for domestic assault are more likely to reoffend than are those who are not punished, which is the reverse of the deterrence theory prediction.

How do we find relevant social theory and prior research? You may already have encountered some of the relevant material in courses pertaining to research questions that interest you, but that won't be enough. The social science research community is large and active, and new research results appear continually in scholarly journals and books. The World Wide Web contains reports on some research even before it is published in journals (like some of the research reviewed in Chapter 1). Conducting a thorough literature review in library sources and checking for recent results on the web are essential steps for evaluating scientific relevance. (See Appendix A for instructions on how to search the literature and the web.)

Journal Link

Read about how researchers apply theory to understand social phenomena.

Theory: A logically interrelated set of propositions about empirical reality.

What Is the Strategy?

When conducting social research, we try to connect theory with empirical data—the evidence we obtain from the real world. Researchers may make this connection in one of two ways:

1. By starting with a social theory and then testing some of its implications with data. This is called **deductive research**; it is most often the strategy used in quantitative methods.
2. By collecting the data and then developing a theory that explains it. This **inductive research** process is typically used with qualitative methods.

A research project can use both deductive and inductive strategies. Let's examine the two different strategies in more detail. We can represent both within what is called the **research circle**.

Deductive Research

In deductive research, we start with a theory and then try to find data that will confirm or deny it. Exhibit 2.2 shows how deductive research starts with a theoretical premise and logically *deduces* a specific expectation. Let's begin with an example of a theoretical idea: When people have emotional and personal connections with coworkers, they will be more committed to their work. We could extend this idea to college life by deducing that if students know their professors well, they will be more engaged in their work. And from this, we can deduce a more specific expectation—or hypothesis—that smaller classes, which allow more student–faculty contact, will lead to higher levels of engagement. Now that we have a hypothesis, we can collect data on levels of engagement in small and large classes and compare them. We can't always directly test the general theory, but we can test specific hypotheses that are deduced from it.

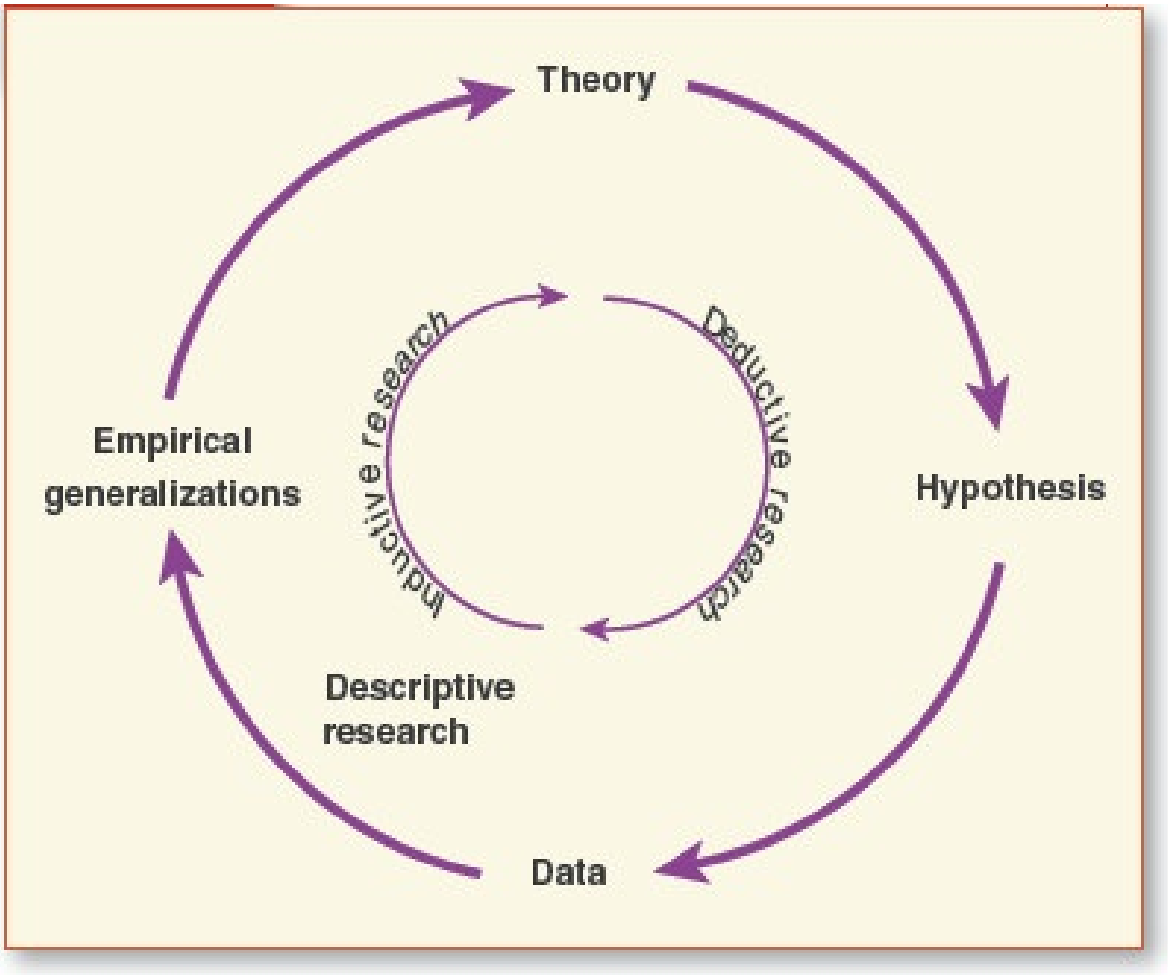
A **hypothesis** states a relationship between two or more **variables**—characteristics or properties that can vary, or change. Classes can be large, like a 400-student introductory psychology course, or they can be small, like an upper-level seminar. Class size is thus a variable. And hours of homework done per week can also vary (obviously); you can do 2 hours or 20. So, too, can engagement vary, as measured in any number of ways. (Nominal designations such as religion are variables, too, because they can vary among Protestant, Catholic, Jew, and so on.)

Interactive Exercises Link

Variables and Hypothesis

But a hypothesis doesn't just state that there is a connection between variables; it suggests that one variable actually influences another—that a change in the first one somehow propels (or predicts, influences, or causes) a change in the second. It says that *if* one thing happens, *then* another thing is likely: *If you* stay up too late, *then* you will be tired the next day. *If* you smoke cigarettes for many years, *then* you are more likely to develop heart disease or cancer. *If* a nation loses a major war, *then* its government is more likely to collapse. And so on.

Exhibit 2.2 The Research Circle



So in a hypothesis, we suggest that one variable influences another—or that the second in some ways "depends" on the first. We may believe, again, that students' reported enthusiasm for a class "depends" on the size of the class. Hence, we call enthusiasm the dependent variable—the variable that *depends* on another, at least partially, for its level. If cigarettes damage your health, then health is the dependent variable; if lost wars destabilize governments, then government stability is the dependent variable.

The predicted result in a hypothesis, then, is called the **dependent variable**. And the hypothesized cause is called the **independent variable** because in the stated hypothesis, it doesn't depend on any other variable. For instance, if we hypothesize that poverty leads to homelessness, then the poverty rate would be the independent variable, and the homeless rate would be the dependent variable.

These terms—*hypothesis*, *variable*, *independent variable*, and *dependent variable*—are used repeatedly in this book and are widely used in all fields of

natural and social science, so they are worth knowing well!

You may have noticed that sometimes an increase in the independent variable leads to a corresponding increase in the dependent variable; in other cases, it leads to a decrease. An increase in your consumption of fatty foods will often lead to a corresponding increase in the cholesterol levels in your blood. But an increase in cigarette consumption leads to a decrease in health. In the first case, we say that the **direction of association** is positive; in the second, we say it is negative. Either way, you can clearly see that a change in one variable leads to a predictable change in the other.

In both explanatory and evaluative research, you should say clearly what you expect to find (your hypothesis) and design your research accordingly to test that hypothesis. Doing this strengthens the confidence we can place in the results. So the deductive researcher (to use a poker analogy) states her expectations in advance, shows her hand, and lets the chips fall where they may. The data are accepted as a fair picture of reality.

Domestic Violence and the Research Circle

The Sherman and Berk (1984) study of domestic violence is a good example of how the research circle works. Sherman and Berk's study was designed to test a hypothesis based on deterrence theory: Arrest for spouse abuse reduces the risk of repeat offenses. In this hypothesis, arrest or release is the independent variable, and variation in the risk of repeat offenses is the dependent variable (it is hypothesized to depend on arrest).

Journal Link

Read about the effects of applying social research on domestic violence.

Sherman and Berk (1984) tested their hypothesis by setting up an experiment in which the police responded to complaints of spouse abuse in one of three ways, one of which was to arrest the offender. When the researchers examined their data (police records for the persons in their experiment), they found that of those arrested for assaulting their spouse, only 13% repeated the offense, compared with a 26% recidivism rate for those who were separated from their spouse by

the police but were not arrested. This pattern in the data, or empirical generalization, was consistent with the hypothesis that the researchers deduced from deterrence theory. The theory thus received support from the experiment (Exhibit 2.3).

Inductive research: The type of research in which general conclusions are drawn from specific data.

Deductive research: The type of research in which a specific expectation is deduced from a general premise and is then tested.

Research circle: A diagram of the elements of the research process, including theories, hypotheses, data collection, and data analysis.

Research in the News

Investigating Child Abuse Doesn’t Reduce It

Congress intended the 1974 Child Abuse Prevention and Treatment Act to increase documentation of and thereby reduce the prevalence of child abuse. However, a review of records of 595 high-risk children nationwide from the ages of 4 to 8 found that those children whose families were investigated were not doing any better than were those whose families were not investigated—except that mothers in investigated families had more depressive symptoms than did mothers in uninvestigated families. Whatever services families were offered after being investigated failed to reduce the risk of future child abuse.

1. What might be the value of a longitudinal design with several surveys of these families?
2. Why might the conclusions have differed if this study had used an experimental design?

News Source: Adapted from Bakalar, Nicholas. 2010. Child abuse investigations didn’t reduce risk, a study finds. *New York Times*, October 12: D3.

Hypothesis: A tentative statement about empirical reality involving a relationship between two or more variables. *Example:* The higher the poverty rate in a community, the higher the percentage of community residents who are homeless.

Variable: A characteristic or property that can vary (take on different values or attributes). *Examples:* poverty rate, percentage of community residents who are homeless.

Dependent variable: A variable that is hypothesized to vary depending on or under the influence of another variable. *Example:* percentage of community residents who are homeless.

Independent variable: A variable that is hypothesized to cause, or lead to, variation in another variable. *Example:* poverty rate.

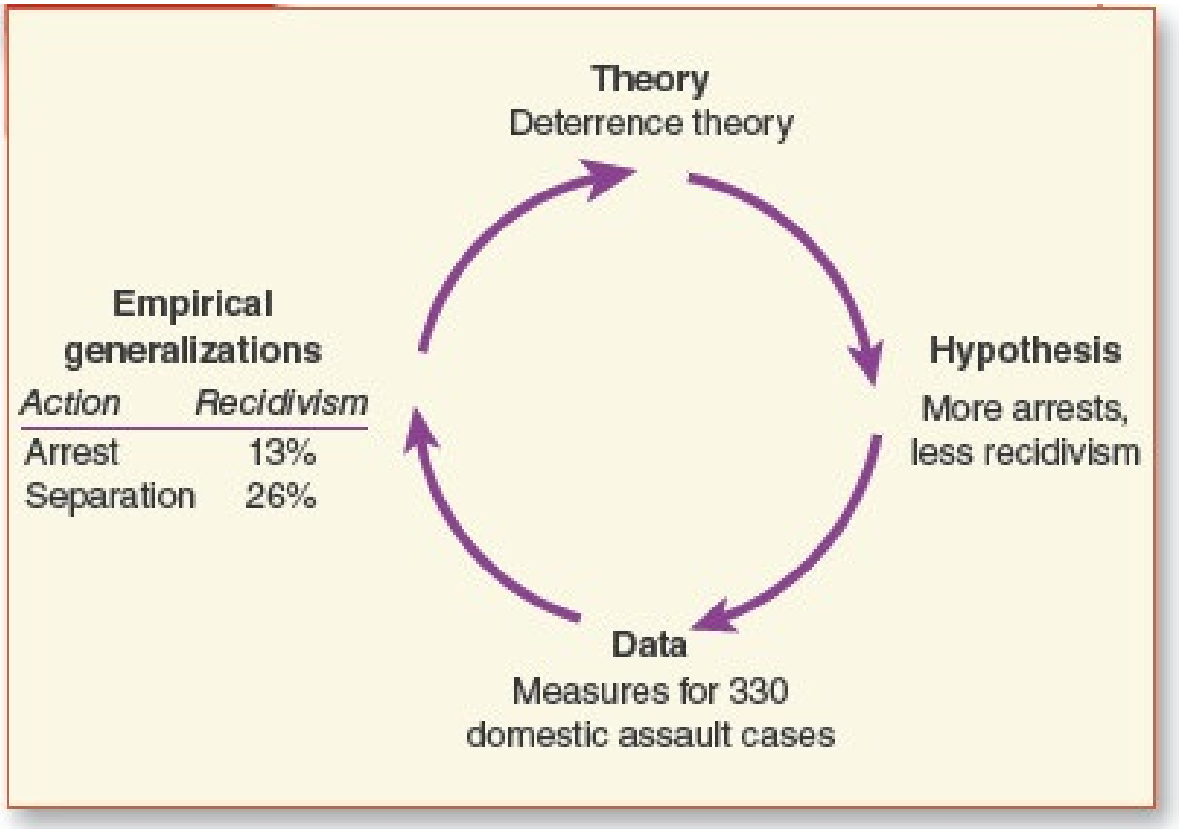
Direction of association: A pattern in a relationship between two variables—that is, the value of a variable tends to change consistently in relation to change in the other variable. The direction of association can be either positive or negative.

Inductive Research

In contrast to deductive research, inductive research begins with specific data, which are then used to develop (*induce*) a theory to account for the data. (Hint: When you start *in* the data, you are doing inductive research.)

One way to think of this process is in terms of the research circle. Rather than starting at the top of the circle with a theory, the inductive researcher starts at the bottom of the circle with data and then moves up to a theory. Some researchers committed to an inductive approach even resist formulating a research question before they begin to collect data. Their technique is to let the question emerge from the social situation itself (Brewer & Hunter 1989: 54–58). In the research for his book *Champions: The Making of Olympic Swimmers*, Dan Chambliss (1988) spent several years living and working with world-class competitive swimmers who were training for the Olympics. Chambliss entered the research with no definite hypotheses and certainly no developed theory about how athletes became successful, what their lives were like, or how they related to their coaches and teams. He simply wanted to understand who these people were, and he decided to report on whatever struck him as most interesting in his research.

Exhibit 2.3 The Research Circle: Minneapolis Domestic Violence Experiment



As it turned out, what Chambliss learned was not how special these athletes were but actually how ordinary they were. Becoming an Olympic athlete was less about innate talent, special techniques, or inspired coaching than it was about actually paying attention to all the little things that make one perform better in one's sport. His theory was *induced* from what he learned in his studies (Chambliss 1988) while being immersed *in* the data.

Research designed using an inductive approach, as in Chambliss's study, can result in new insights and provocative questions. **Inductive reasoning** also enters into deductive research when we find unexpected patterns in data collected for testing a hypothesis. Sometimes such patterns are **anomalous**, in that they don't seem to fit the theory being proposed, and they can be **serendipitous**, in that we may learn exciting, surprising new things from them. Even if we do learn inductively from such research, the adequacy of an explanation formulated after the fact is necessarily less certain than an explanation presented before the collection of data. Every phenomenon can always be explained in some way. Inductive explanations are more trustworthy if they are tested subsequently with deductive research. Great insights and ideas can come from inductive studies, but verifiable proof comes from deductive

research.

Journal Link

Read about inductive and deductive research techniques in the wake of another disaster.

An Inductive Study of Response to a Disaster

Qualitative research is often inductive: To begin, the researcher observes social interaction or interviews social actors in depth, and then develops an explanation for what has been found. The researchers often ask such questions as these: What is going on here? How do people interpret these experiences? Why do people do what they do? Rather than testing a hypothesis, the researchers try to make sense of some social phenomenon.

In 1972, for example, towns along the 17-mile Buffalo Creek hollow in West Virginia were wiped out when a dam at the top of a hollow broke, sending 132 million gallons of water, mud, and garbage crashing down through the towns that bordered the creek. After the disaster, sociologist Kai Erikson went to the Buffalo Creek area and interviewed survivors. In the resulting book, *Everything in Its Path*, Erikson (1976) described the trauma suffered by those who survived the disaster. His explanation of their psychological destruction—an explanation that grew out of his interviews with the residents—was that people were traumatized not only by the violence of what had occurred but also by the "destruction of community" that ensued during the recovery efforts. Families were transplanted all over the area with no regard for placing them next to their former neighbors. Extended families were broken up in much the same way, as federal emergency housing authorities relocated people with little concern for whether they knew the people with whom they would be housed. Church congregations were scattered, lifelong friends were resettled miles apart, and entire neighborhoods simply vanished, both physically—that is, their houses were destroyed—and socially. Erikson's explanation grew out of his in-depth immersion in his data—the conversations he had with the people themselves.

Inductive explanations such as Erikson's feel authentic because we hear what people have to say in their own words and we see the social world as they see it.

These explanations are often richer and more finely textured than are those in deductive research; however, they are probably based on fewer cases and drawn from a more limited area.

Research That Matters

The Sherman and Berk domestic violence study did not, however, end the debate about how best to respond to incidents. By the 1990s, the Charlotte-Mecklenburg (North Carolina) Police Department (CMPD) had been responding to reports of violence against intimate partners by arresting many of the suspects. Unfortunately, six months after the arrests, it appeared that suspects who had been arrested were just as likely to reoffend, as were those who were simply advised to cool off. In 1995, the CMPD decided to try a different approach to domestic violence cases. CMPD developed a special domestic violence unit that took a comprehensive team approach to investigating cases and assisting victims. Professors M. Lyn Exum, Jennifer L. Hartman, Paul C. Friday, and Vivian B. Lord, at the University of North Carolina in Charlotte, set out to see if this approach worked. They checked the arrest records of 891 domestic violence cases to see if suspects processed through the special unit were less likely to reoffend than were those who were processed with standard police practices. Exum and her colleagues found that 29.3% of the suspects processed by the domestic violence unit reoffended, compared with 36.9% of those processed by a standard police patrol unit. There was a little, but not much, difference.

Source: Adapted from Exum, M. Lyn, Jennifer L. Hartman, Paul C. Friday, and Vivian B. Lord. 2010. Policing domestic violence in the post-SARP era: The impact of a domestic violence police unit. *Crime & Delinquency* 20(10): 1–34.

Inductive reasoning: The type of reasoning that moves from the specific to the general.

Anomalous: Unexpected patterns in data that do not seem to fit the theory being proposed.

Serendipitous: Unexpected patterns in data, which stimulate new ideas or theoretical approaches.

Descriptive Research: A Necessary Step

Both deductive and inductive research move halfway around the research circle, connecting theory with data. Descriptive research does not go that far, but it is still part of the research circle shown earlier in Exhibit 2.2. Descriptive research starts with data and proceeds only to the stage of making empirical generalizations; it does not generate entire theories.

Research|Social Impact Link

Learn about a study that uses descriptive research.

Valid description is actually critical in all research. The Minneapolis Domestic Violence Experiment was motivated partly by a growing body of descriptive research indicating that spouse abuse is very common: 572,000 reported cases of women victimized by a violent partner each year; 1.5 million women (and 500,000 men) requiring medical attention each year from a domestic assault (Buzawa & Buzawa 1996: 1–3).

Much important research for the government and private organizations is primarily descriptive: How many poor people live in this community? Is the health of the elderly improving? How frequently do convicted criminals return to crime? Description of social phenomena can stimulate more ambitious deductive and inductive research. Simply put, good description of data is the cornerstone for the scientific research process and an essential component of understanding the social world.

What Is the Design?

Researchers usually start with a question, although some begin with a theory or a strategy. If you're very systematic, the *question* is related to a *theory*, and an appropriate *strategy* is chosen for the research. All of these, you will notice, are critical defining issues for the researcher. If your research question is trivial (How many shoes are in my closet?), or your theory sloppy (More shoes reflect better fashion sense), or your strategy inappropriate (I'll look at lots of shoes and see what I learn), the project is doomed from the start.

But let's say you've settled these first three elements of a sound research study. Now we must begin a more technical phase of the research: the design of a study. From this point on, we will be introducing a number of terms and definitions that may seem arcane or difficult. In every case, though, these terms will help you clarify your thinking. Like exact formulae in an algebra problem or precisely the right word in an essay, these technical terms help, or even require, scientists to be absolutely clear about what they are thinking—and to be precise in describing their work to other people.

An overall research strategy can be implemented through several different types of research design. One important distinction between research designs is whether data are collected at one point in time—a **cross-sectional research design**—or at two or more points in time—a **longitudinal research design**. Another important distinction is between research designs that focus on individuals—the **individual unit of analysis**—and those that focus on groups, or aggregates of individuals—the **group unit of analysis**.

Cross-Sectional Designs

In a cross-sectional design, all of the data are collected at one point in time. In effect, you take a *cross-section*—a slice that cuts across an entire population—and use that to see all the different parts, or sections, of that population. Imagine cutting out a slice of a tree trunk, from bark to core. In looking at this cross-section, one can see all the different parts, including the rings of the tree. In social research, you might do a cross-sectional study of a college's student body, with a sample that includes freshmen through seniors. This "slice" of the population, taken at a single point in time, allows one to compare the different groups.

Research|Social Impact Link

Read an article that uses evidence from multiple cross-sectional studies.

But cross-sectional studies, because they use data collected at only one time, suffer from a serious weakness: They don't directly measure the impact of time. For instance, you may see that seniors at your college write more clearly than do freshmen. You might conclude, then, that the difference is because of what transpired over time, that is, what they learned in college. But it might actually be because this year's seniors were recruited under a policy that favored better writers. In other words, the cross-sectional study doesn't distinguish if the seniors have learned a lot in college or if they were just better than this year's freshmen when they first enrolled.

Or let's say that in 2015, you conduct a study of the U.S. workforce and find that older workers make more money than younger workers do. You may conclude (erroneously) that as one gets older, one makes more money. But you didn't actually observe that happening because you didn't track actual people over time. It *may* be that the older generation (say, people born in 1965) have just enjoyed higher wages all along than have people born in 1985.

With a cross-sectional study, we can't be sure which explanation is correct, and that's a big weakness. Of course, we could ask workers what they made when they first started working, or we could ask college seniors what test scores they received when they were freshmen, but we are then injecting a *longitudinal*

element into our cross-sectional research design. Because of the fallibility of memory and the incentives for distorting the past, taking such an approach is not a good way to study change over time.

Cross-sectional research design: A study in which data are collected at only one point in time.

Longitudinal research design: A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.

Individual unit of analysis: A unit of analysis in which individuals are the source of data and the focus of conclusions.

Group unit of analysis: A unit of analysis in which groups are the source of data and the focus of conclusions.

Longitudinal Designs

In longitudinal research, data are collected over time. By measuring independent and dependent variables at each of several different times, the researcher can determine whether change in the independent variable actually precedes change in the dependent variable—that is, whether the hypothesized cause comes before the effect, as a true cause must. In a cross-sectional study, when the data are all collected at one time, you can't really show if the hypothesized cause occurs first; in longitudinal studies, though, you can see if a cause occurs and then, later in time, an effect occurs. So if possible to do, longitudinal research is always preferable.

But collecting data more than once takes time and work. Often researchers simply cannot, or are unwilling to, delay completion of a study for even 1 year to collect follow-up data. In student research projects, longitudinal research is typically not possible because you have to finish up the project quickly. Still, many research questions really should have a long follow-up period: What is the impact of job training on subsequent employment? How effective is a school-based program in improving parenting skills? Under what conditions do traumatic experiences in childhood result in later mental illness? The value of longitudinal data is great, so every effort should be made to develop longitudinal research designs whenever they are appropriate.

Three basic research designs are shown in Exhibit 2.4. The first is a simple *cross-sectional* design; it is not longitudinal.

Audio Link

Listen to more information on polls.

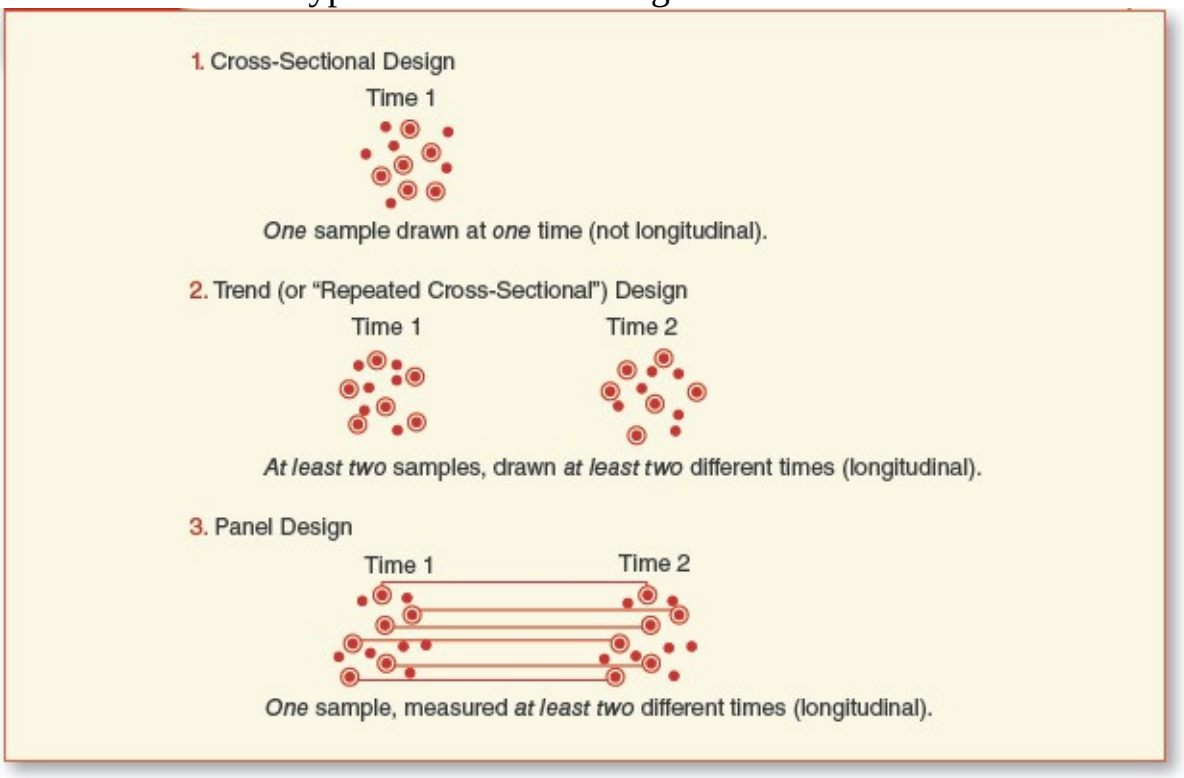
The second is a cross-sectional study that is then *repeated* at least once; therefore, this approach is referred to as a *repeated cross-sectional* or a *trend* design because it can capture trends over time; it is longitudinal. The frequency of the follow-up measurements can vary, ranging from a simple before-and-after design with just one follow-up to studies in which various indicators are measured every month for many years. In such trend studies, members of the

sample are rotated or completely replaced each time a measurement is done.

The third design, also longitudinal, is called a *panel* study. A panel study uses a single sample that is studied at multiple points across time; the same people, for instance, will be asked questions on multiple occasions, so researchers can learn how they change and develop as individuals.

Let's consider the longitudinal designs to see how they are done and what are their strengths and weaknesses.

Exhibit 2.4 Three Types of Research Designs


1. Cross-Sectional Design
Time 1
One sample drawn at one time (not longitudinal).
2. Trend (or "Repeated Cross-Sectional") Design
Time 1
Time 2
At least two samples, drawn at least two different times (longitudinal).
3. Panel Design
Time 1
Time 2
One sample, measured at least two different times (longitudinal).

Trend Designs

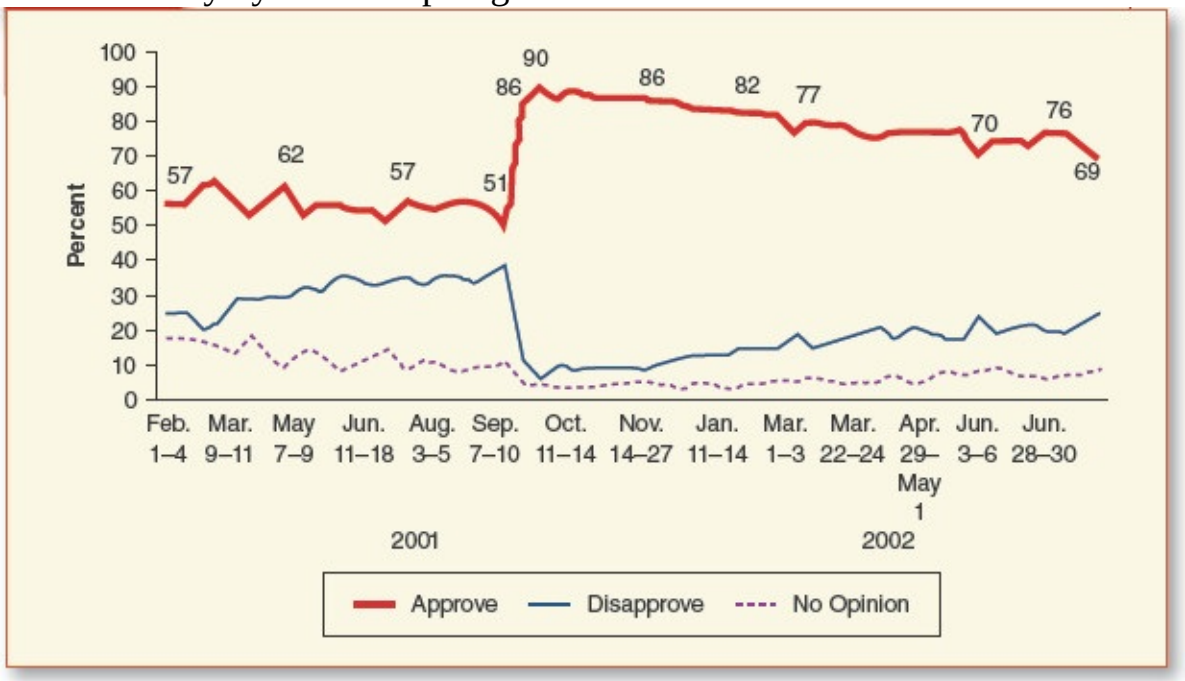
Trend designs, also known as **repeated cross-sectional studies**, are conducted as follows:

1. A sample is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some people leave the population and others enter it.
3. At Time 2, a different sample is drawn from this population.

The Gallup polls, begun in the 1930s, are a well-known example of trend studies. One Gallup poll, for instance, asks people how well they believe the U.S. president is doing his job (Exhibit 2.5). Every so often, the Gallup organization takes a sample of the U.S. population (usually about 1,400 people) and asks them this question. Each time, Gallup researchers ask a different, though roughly demographically equivalent, group of people the question; they aren't talking to the same people every time. Then they use the results of a series of these questions to analyze trends in support for presidents. That is, they can see when support for presidents is high and when it is low, in general. This is a trend study. Exhibit 2.5 shows the dramatic change in the public's approval rating of President George W. Bush following the September 11, 2001, bombings.

When the goal is to determine whether a population (not necessarily individuals) has changed over time, trend (or "repeated cross-sectional") designs are appropriate. Has support for gay marriage increased among Americans in the past 20 years? Are employers more likely to pay maternity benefits today than they were in the 1950s? Are college students today more involved in their communities than college students were 10 years ago? These questions concern changes in populations as a whole, not changes in individuals.

Exhibit 2.5 George W. Bush Approval Ratings, Before and After Sept. 11, 2001: A Trend Study by the Gallup Organization



Source: The Gallup Organization. August 20, 2002. Poll Analyses, July 29, 2002. Bush Job Approval Update.

Trend (repeated cross-sectional) design: A longitudinal study in which data are collected at two or more points in time from different samples of the same population.

Panel Designs

When we need to know whether specific individuals in a population have changed, we must turn to a **panel design**. For their book *How College Works* (2014), Dan Chambliss and Chris Takacs selected a panel of 100 random students entering college in 2001. Then each student was interviewed once a year for each of their 4 years in college; then they were interviewed every 2 years after graduation until 2010. The goal was to determine which experiences in their college career were valuable and which were a hindrance to their education. By following the same people over time, we can see how changes happen in the lives of individual students.

Panel designs allow clear identification of changes in the units (individuals, groups, or whatever) we are studying. Here is the process for conducting fixed-sample panel studies:

1. A sample (called a *panel*) is drawn from a population at Time 1, and data are collected from the sample (for instance, 100 freshmen are selected and interviewed).
2. As time passes, some panel members become unavailable for follow-up, and the population changes (some students transfer to other colleges or decline to continue participating).
3. At Time 2, data are collected (the remaining students are reinterviewed) from the same people (the panel) as at Time 1, except for those people who cannot be located.

A panel design allows us to determine how individuals change, as well as how the population as a whole has changed; this is a great advantage. However, panel designs are difficult to implement successfully and often are not even attempted, for two reasons:

1. *Expense and attrition*—It can be difficult and expensive to keep track of individuals over a long period, and inevitably the proportion of panel members who can be located for follow-up will decline over time. Panel studies often lose more than one quarter of their members through attrition (Miller 1991: 170).
2. *Subject fatigue*—Panel members may grow weary of repeated interviews and drop out of the study, or they may become so used to answering the standard questions in the survey that they start giving stock answers rather than actually thinking about their current feelings or actions (Campbell 1992). This is called the problem of *subject fatigue.*

Although quite difficult to do, panel studies can be scientifically quite valuable and intrinsically fascinating. In the British "Up" documentary film series, a group of 14 British 7-year-olds were filmed in 1964, for a movie called "7 Up". Every 7 years since then, the same people have been interviewed, for what has become one of the most extraordinary documentaries ever made. The latest movie is called "56 Up, " and shows the current lives of the same people, now much older. Only one has dropped out completely. The series as a whole thus follows these people through their lives, and is immensely revealing of, for instance, the ways their social class has affected them.

Panel design: A longitudinal study in which data are collected from the same individuals—the panel—at two or more points in time.

Cohort Designs

Among other uses, longitudinal studies can be designed to track the results of either an event (such as the 9/11 bombings, or the 2008 economic crash) or the progress of a specific historical generation (for instance, people born in 1996). In these cases, the specific group of people being studied is known as a **cohort**, and the study is using a **cohort design**. If you were doing a trend study, for instance, the cohort would be the population from which you draw your series of samples. If you were doing a panel study, the cohort provides the population from which the panel itself is drawn. Examples of cohorts include the following:

- *Birth cohorts*—those who share a common period of birth—for example, "baby boomers" born after World War II, "millennials" who became adults around 2000, "digital natives" born since the Internet became pervasive,

and so forth.

- *Seniority cohorts*—those who have worked at the same place for about 5 years, about 10 years, and so on.
- *Event cohort*—people who have shared an event, for instance, all the victims of Hurricane Sandy that hit the Northeast coast of the United States in 2012. Many panel studies are based on cohorts because the people selected by definition all start in the research at the same specific time in history; the researcher needs to be aware that their cohort status (when they are living) may affect the results.

Video Link

Watch a video about cohort designs.

We can see the value of longitudinal research using a cohort design in comparing two studies that estimated the impact of public and private schooling on high school students' achievement test scores. In an initial cross-sectional (not longitudinal) study, James Coleman, Thomas Hoffer, and Sally Kilgore (1982) compared standardized achievement test scores of high school sophomores and seniors in public, Catholic, and other private schools. The researchers found that test scores were higher in the private (including Catholic) high schools than in the public high schools.

But was this difference a causal effect of private schooling? Perhaps the parents of higher-performing children were choosing to send them to private schools rather than to public ones. So Coleman and Hoffer (1987) went back to the high schools and studied the test scores of the former sophomores 2 years later, when they were seniors; in other words, the researchers used a panel (longitudinal) design. This time, they found that the verbal and math achievement test scores of the Catholic school students had increased more over the 2 years than the scores of the public school students had. Irrespective of students' initial achievement test scores, the Catholic schools seemed to "do more" for their students than did the public schools. The researchers' causal conclusion rested on much stronger ground because they used a longitudinal panel design.

Cohort: Individuals or groups with a common starting point.

Cohort design: A longitudinal study in which data are collected at two or more points in time from individuals in a cohort.

Units and Levels of Analysis

Units of analysis are the things you are studying, whose behavior you want to understand. Often these are individual people (e.g., why do certain students work harder?), but they can also be, for instance, families, groups, colleges, governments, or nations. All of these could be units of analysis for your research. Sociologist Erving Goffman, writing about face-to-face interaction, became famous partly because he realized that the interaction itself—not just the people in it—could be a unit of analysis. Goffman argued that interactions as such worked in certain ways, apart from the individuals who happened to be joining them: "Not, then, men and their moments. Rather, moments and their men" (Goffman 1967: 3). Researchers must always be clear about what is the level of social life they are studying: What are their units of analysis? The units of analysis are the entities you are studying and trying to learn about.

As the examples suggest, units exist at different *levels* of collectivity, from the most micro (small) to the most macro (large). Individual people are easily seen and talked to, and you can learn about them quite directly. A university, however, although you can certainly visit it and walk around it, is harder to visualize, and data regarding it may take longer to gather. Finally, a nation is not really a "thing" at all and can never be seen by human eyes; understanding such a unit may require many years of study. People, universities, and nations exist at different *levels* of social reality. And as probably already known, groups don't act like individuals do.

Sometimes researchers confuse levels of analysis, mistakenly using data from one level to draw conclusions about a different level. Even the best social scientists fall into this trap. In Émile Durkheim's classic (1951) study of suicide, for example, nationwide suicide rates were compared for Catholic and Protestant countries (in an early stage of his research). The data on suicide were collected for individual people, and religion was tallied for individuals as well. Then Durkheim used aggregated numbers to characterize entire countries as being high or low suicide countries and as Protestant (England, Germany, Norway) or Catholic (Italy, France, Spain) countries. He found that Catholic countries had lower rates of suicide than Protestant countries had. His accurate finding was about countries, then, not about people; the unit of analysis was the country, and he ranked countries by their suicide rates. Yes, the data were collected from

individuals and were about individuals, but it had been combined (aggregated) to describe entire nations. Thus, Durkheim's units of analysis were countries. So far, so good.

But Durkheim then made his big mistake. He used his findings from one level of analysis to make statements about units at a different level. He used country data to draw conclusions about individuals, claiming that Catholic individuals were less likely than were Protestant individuals to commit suicide. Much of his later discussion in *Suicide* (1951) was about why Catholic individuals would be less likely to kill themselves. Perhaps they are, but we can't be sure based on aggregate data. It could be that Protestant individuals were more likely to kill themselves in Durkheim's time when they lived in areas with high numbers of Catholics.

Encyclopedia Link

Read more details on levels of analysis.

Confusions about levels of analysis can take several forms (Lieberson 1985). Durkheim's mistake was to use findings from a "higher" level (countries) to draw conclusions about a "lower" level (individuals). This is called the **ecological fallacy** because the *ecology*—the broader surrounding setting, in this case a country—is mistakenly believed to straightforwardly model how individuals will act as well. The ecological fallacy occurs when group-level data are used to draw conclusions about individual-level processes. It's a mistake, and a common one.

Try to spot the ecological fallacy in each of the following deductions. The first half of each sentence is true, but the second half doesn't logically follow from the first:

- Richer countries have higher rates of heart disease; therefore, richer people have higher rates of heart disease.
- Florida counties with the largest number of black residents have the highest rates of Ku Klux Klan membership; therefore, blacks join the Klan more than whites.
- In the 2012 presidential election, Republicans won the House of

Representatives, but Democrats held onto the Senate, and President Obama was reelected; therefore, Americans want a divided government.

In each case, a group-level finding from data is used to draw (erroneous) conclusions about individuals. In rich countries, yes, there is more heart disease, but actually, it's among the poor individuals within those countries. Florida counties with more black people attract more white individuals to the Klan. And although the United States (as a whole) was certainly divided in the 2012 election, just as certainly many individual Americans, both Republican and Democratic, had no ambivalence whatsoever about who were their favorite candidates. *America* as a whole may "want a divided government," but relatively few *Americans* do.

Audio Link

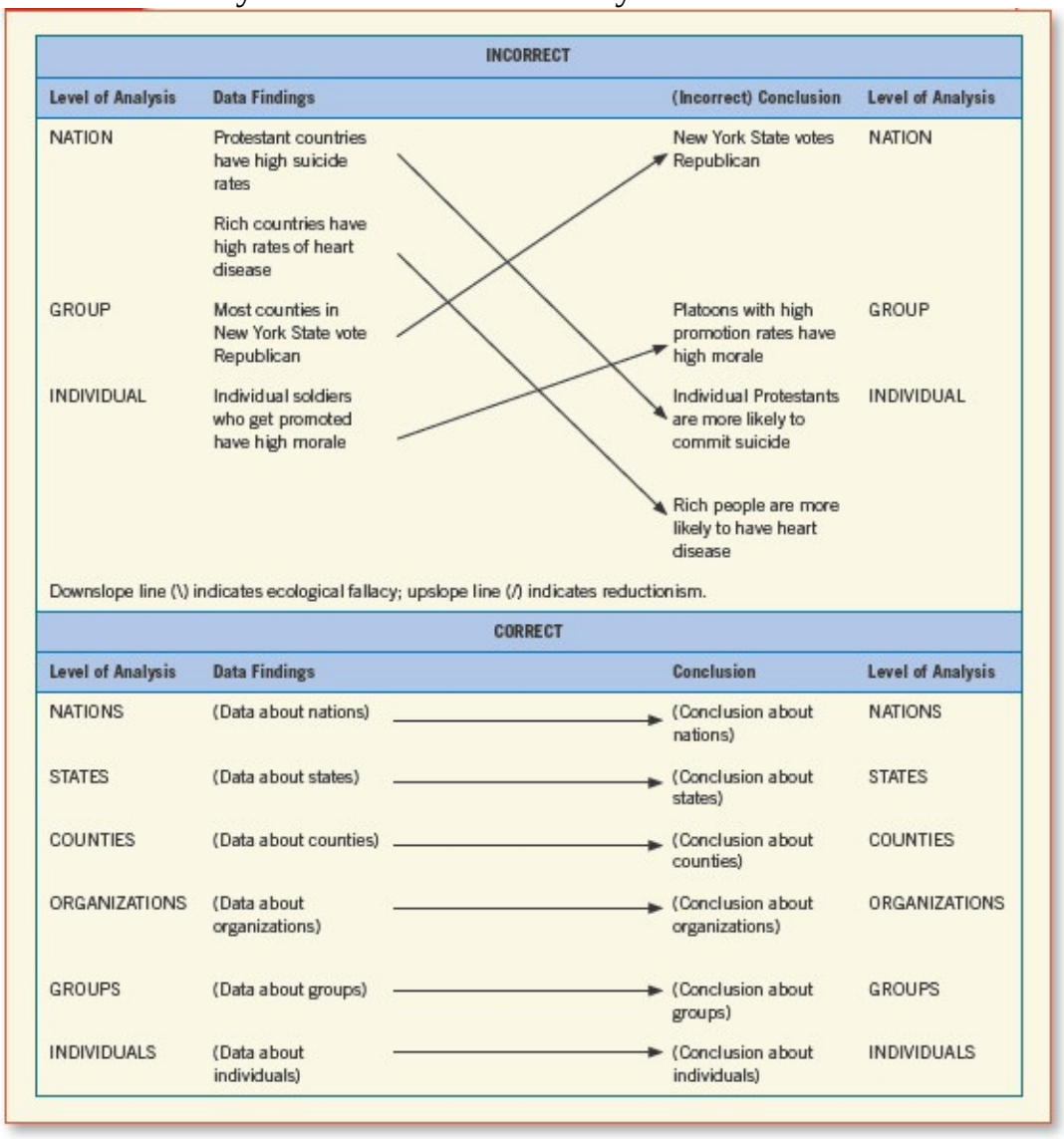
Listen to a study that involves ecological fallacy.

A researcher who draws such hasty conclusions about individual-level processes from group-level data is committing an ecological fallacy. In August 2006, the *American Sociological Review* published a fierce exchange in which Mitchell Duneier, a well-known field researcher from Princeton University, attacked a very popular book *Heat Wave*, by Eric Klinenberg. *Heat Wave* vividly described how hundreds of poor people in Chicago died during a heat wave in July 1995. Klinenberg argued that the deaths were the result of deteriorating community conditions—for instance, that vulnerable old people, afraid to go outside and possibly be attacked or mugged, remained indoors despite literally killing temperatures in their homes. Duneier (2006) claimed that Klinenberg lacked any data on individual deaths to show that this is what happened, although it was clear that community conditions mattered. But, Duneier argued, the fact that certain features prevailed in the stricken communities did not mean that it was those conditions themselves that led to individual deaths. Klinenberg (2006) disagreed, strongly.

So, conclusions about processes at the individual level must be based on individual-level data; conclusions about group-level processes must be based on data collected about groups (Exhibit 2.6.)

Exhibit 2.6 Levels of analysis. Data from one level of analysis should lead to

conclusions only about that level of analysis.



INCORRECT

Level of Analysis	Data Findings	(Incorrect) Conclusion	Level of Analysis
NATION	Protestant countries have high suicide rates	New York State votes Republican	NATION
	Rich countries have high rates of heart disease		
GROUP	Most counties in New York State vote Republican	Platoons with high promotion rates have high morale	GROUP
INDIVIDUAL	Individual soldiers who get promoted have high morale	Individual Protestants are more likely to commit suicide	INDIVIDUAL
		Rich people are more likely to have heart disease	

Downslope line (\) indicates ecological fallacy; upslope line (/) indicates reductionism.

CORRECT

Level of Analysis	Data Findings	Conclusion	Level of Analysis
NATIONS	(Data about nations)	(Conclusion about nations)	NATIONS
STATES	(Data about states)	(Conclusion about states)	STATES
COUNTIES	(Data about counties)	(Conclusion about counties)	COUNTIES
ORGANIZATIONS	(Data about organizations)	(Conclusion about organizations)	ORGANIZATIONS
GROUPS	(Data about groups)	(Conclusion about groups)	GROUPS
INDIVIDUALS	(Data about individuals)	(Conclusion about individuals)	INDIVIDUALS

We don’t want to leave you with the belief that conclusions about individual processes based on group-level data are *necessarily* wrong. We just don’t know for sure. Suppose, for example, that we find that communities with higher average incomes have lower crime rates. Perhaps something about affluence improves community life such that crime is reduced; that’s possible. Or, it may be that the only thing special about these communities is that they have more individuals with higher incomes, who tend to commit fewer crimes. Even though we collected data at the group level and analyzed them at the group level, they may reflect a causal process at the individual level (Sampson & Lauritsen 1994: 80–83). The ecological fallacy just reminds us that we can’t *know* about individuals without having individual-level information.

Confusion between levels of analysis also occurs in the other direction, when data from the individual level are used to draw conclusions about group behavior. For instance, you may know the personal preferences of everyone on a hiring committee, so you try to predict whom the committee will decide to hire, but you could easily be wrong. Or you may know two good individuals who are getting married, so you think that the marriage (the higher-level unit) will be good, too. But often, such predictions are wrong because groups as units don't work like individuals. Nations often go to war even when most of their people (individually) don't want to. Adam Smith, in the 1700s, famously pointed out that millions of people (individuals) acting selfishly could in fact produce an economy (a group) that acted selflessly, helping almost everyone. You can't predict higher-level processes or outcomes from lower-level ones. You can't, in short, always reduce group behavior to individual behavior added up; doing so is called the **reductionist fallacy**, or **reductionism** (because it *reduces* group behavior to that of individuals), and it's basically the reverse of the ecological fallacy.

Both involve confusion of levels of analysis.

Units of analysis: The entities being studied, whose behavior is to be understood.

Ecological fallacy: An error in reasoning in which conclusions about individual-level processes are drawn from group-level data.

Careers and Research

Russell K. Schutt, PhD

Source: Russell K. Schutt

Congratulations! You can now take the first step to becoming a social researcher and a consumer of social research, by developing a research question and deciding to begin the process of research. As a result, I hope you are beginning to see the potential for using social research methods to understand issues that matter to you, to identify policies that can help others, and to add to the body of social science knowledge.

There are many ways to develop research interests, and I'd like to share with you some of my own experiences about that. My research experience as a graduate student at the University of Illinois at Chicago and as a postdoctoral fellow at Yale University was in the sociology of organizations, occupations, and law. My interest in research in a new area, homelessness and mental health, developed gradually in subsequent years after I joined the faculty at the University of Massachusetts Boston. One day, I found in my mailbox a plea from a recent graduate for help with "computerizing the case management records" at the shelter for which she had started to work. I was scheduled to teach a graduate course in computer applications and decided to take on this effort as a class project.

With the experience my students and I gained in the project, I was able to write a proposal with a colleague for funding from a new university initiative in health research. As a result of making connections with other researchers and service providers, writing research reports for funders, reading the relevant research literature, and investigating the needs of homeless persons, I was able to write additional research proposals to study homeless persons and shelter services that

were funded by the university and by local service programs. Although my proposal to the National Science Foundation with medical sociologist Mary Fennell to study organizational change in shelters was not funded, the pilot study we carried out led to an invitation to join a team of researchers who were responding to a special National Institute of Mental Health (NIMH) request for proposals about housing and services for homeless persons with serious mental health problems.

The $13.1 million our team received from NIMH and Housing and Urban Development (HUD) allowed us to carry out a longitudinal randomized test of the value of group and independent housing using a mixed-method design that in turn led to many journal articles, some book chapters, and one book. From the small beginning of a class project involving secondary data analysis, to cross-sectional surveys of homeless persons, shelter staff, and shelter directors, to longitudinal evaluation research and then a randomized experiment, these research projects became increasingly sophisticated and supported more significant contributions to social policy and the scholarly literature.

So be prepared to follow your interests, take advantage of opportunities, and maintain ambitious goals!

Reductionist fallacy (reductionism): An error in reasoning that occurs when incorrect conclusions about group-level processes are based on individual-level data.

Conclusion

Social researchers can find many questions to study, but not all questions are equally worthy. The ones that warrant the expense and effort of social research are feasible, socially important, and scientifically relevant.

Selecting a worthy research question does not guarantee a worthwhile research project. The simplicity of the research circle presented in this chapter belies the complexity of the social research process. In the following chapters, we will focus on particular aspects of that process. Chapter 4 examines the interrelated processes of conceptualization and measurement, arguably the most important parts of research. Measurement validity is the foundation for the other two aspects of validity, which are discussed in Chapters 5 and 6. Chapter 5 reviews the meaning of generalizability and the sampling strategies that help us to achieve this goal. Chapter 6 introduces the third aspect of validity—causal validity—and illustrates different methods for achieving causal validity and explains basic experimental data collection. Chapters 7 and 9 introduce approaches to data collection—surveys and qualitative research—that help us, in different ways, to achieve validity.

You are now forewarned about the difficulties that all scientists, but social scientists in particular, face in their work. We hope that you will return often to this chapter as you read the subsequent chapters, when you criticize the research literature, and when you design your own research projects. To be conscientious, thoughtful, and responsible—this is the mandate of every social scientist. If you formulate a feasible research problem, ask the right questions in advance, try to adhere to the research guidelines, and steer clear of the most common difficulties, you will be well along the road to fulfilling this mandate.

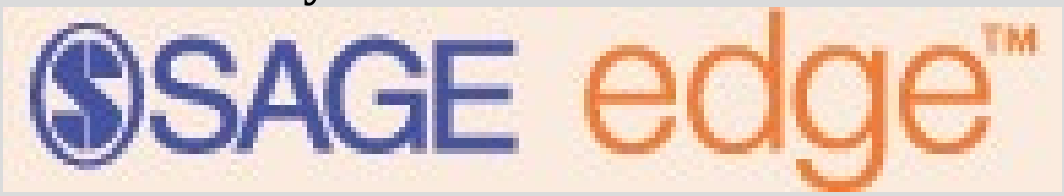
Key Terms

Anomalous 27
Cohort 32
Cohort design 32
Cross-sectional research design 29
Deductive research 24
Dependent variable 26
Direction of association 26
Ecological fallacy 34
Group unit of analysis 29
Hypothesis 25
Independent variable 26
Individual unit of analysis 29
Inductive reasoning 27
Inductive research 24
Longitudinal research design 29
Panel design 32
Reductionist fallacy (reductionism) 37
Research circle 24
Serendipitous 27
Social research question 22
Theory 23
Trend (repeated cross-sectional) design 31
Units of analysis 33
Variable 25

Highlights

- Research questions should be feasible (within the time and resources available), socially important, and scientifically relevant.
- Building social theory is a major objective of social science research. Investigate relevant theories before starting social research projects, and draw out the theoretical implications of research findings.
- The type of reasoning in most research can be described as primarily deductive or primarily inductive. Research based on deductive reasoning proceeds from general ideas, deduces specific expectations from these ideas, and then tests the ideas with empirical data. Research based on inductive reasoning begins with (*in*) specific data and then develops (*induces*) general ideas or theories to explain patterns in the data.
- It may be possible to explain unanticipated research findings after the fact, but such explanations have less credibility than those that have been tested with data collected for the purpose of the study.
- The scientific process can be represented as circular, with connections from theory, to hypotheses, to data, and to empirical generalizations. Research investigations may begin at different points along the research circle and traverse different portions of it. Deductive research begins at the point of theory; inductive research begins with data but ends with theory. Descriptive research begins with data and ends with empirical generalizations.
- Research designs vary in their units of analysis—the primary distinctions being individual or group—and in their collection of data at one point in time—a cross-sectional design—or at two or more points in time—a longitudinal design, with three options: a trend design, a panel design, or a cohort design.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Pick a social issue about which you think research is needed. Draft three research questions about this issue. Refine one of the questions and evaluate it in terms of the three criteria for good research questions.
2. Identify variables that are relevant to your three research questions. Now formulate three related hypotheses. Which are the independent and which are the dependent variables in these hypotheses?
3. If you were to design research about domestic violence, would you prefer an inductive approach or a deductive approach? Explain your preference. What would be the advantages and disadvantages of each approach? Consider in your answer the role of social theory, the value of searching the literature, and the goals of your research.
4. Sherman and Berk's (1984) study of the police response to domestic violence tested a prediction derived from deterrence theory. Propose hypotheses about the response to domestic violence that are consistent with labeling theory. Which theory seems to you to provide the best framework for understanding domestic violence and how to respond to it?
5. Review our description of the research projects in the section "Social Research in Practice" in Chapter 1. Can you identify the stages of each project corresponding to the points on the research circle? Did each project include each of the four stages? Which theory (or theories) seem applicable to each of these projects? What were the units of analysis? Were the designs cross-sectional or longitudinal?

Finding Research

1. State a problem for research—some feature of social life that interests you. If you have not already identified a problem for study, or if you need to evaluate whether your research problem is doable, a few suggestions should help to get the ball rolling and keep you on course.
    1. Jot down several questions that have puzzled you about people and social relations, perhaps questions that have come to mind while reading textbooks or research articles, talking with friends, or hearing news stories.
    2. Now take stock of your interests, your opportunities, and the work of others. Which of your research questions no longer seem feasible or interesting? What additional research questions come to mind? Pick out one question that is of interest and seems feasible and that has probably been studied before.
    3. Do you think your motives for doing the research would affect how the research is done? How? Imagine several different motives for doing the research. Might any of them affect the quality of your research? How?
    4. Write out your research question in one sentence; then elaborate on it in one paragraph. List at least three reasons why it is a good research question for you to investigate. Then present your question to your classmates and instructor for discussion and feedback.

2. Review Appendix A: Finding Information, and then search the literature (and the Internet) on the research question you identified. Copy down at least five citations for articles (with abstracts from Cambridge Scientific Abstracts [CSA] Sociological Abstracts) and two websites reporting research that seems highly relevant to your research question. Look up at least two of these articles and one of the websites. Inspect the article bibliographies and the links at the website, and identify at least one more relevant article and website from each source.

   Write a brief description of each article and website you consulted and evaluate its relevance to your research question. What additions or changes to your thoughts about the research question do the sources suggest?
3. To brush up on a range of social theorists, visit the site www.sociologyprofessor.com, pick a theorist, and read some of what you find. What social phenomena does this theorist focus on? What hypotheses seem consistent with his or her theorizing? Describe a hypothetical research project to test one of these hypotheses.
4. You've been assigned to write a paper on domestic violence and the law. To start, you can review relevant research on the American Bar Association's website (www.americanbar.org/groups/domestic_violence/resources/statistics.html). What does the research summarized at this site suggest about the prevalence of domestic violence, its distribution about social groups, and its causes and effects? Write your answers in a one- to two-page report.

Critiquing Research

1. Using recent newspapers or magazines, find three articles that report on large interview or survey research studies. Describe each study briefly. Then say (a) whether the study design was longitudinal or cross-sectional and (b) if that mattered—that is, if the study's findings would possibly have been different using the alternative design.
2. Search the journal literature for three studies concerning some social program or organizational policy after you review the procedures in Appendix A. Several possibilities are research on Head Start, on the effects of welfare payments, on boot camps for offenders, and on standardized statewide testing in the public schools. Would you characterize the findings as largely consistent or inconsistent? How would you explain discrepant findings?

Doing Research

1. Formulate four research questions about support for capital punishment. Provide one question for each research purpose: descriptive, exploratory, explanatory, and evaluative.
2. State four hypotheses in which support for capital punishment is the dependent variable and some other variable is the independent variable.
    1. Justify each hypothesis in a sentence or two.
    2. Propose a design to test each hypothesis. Design the studies to use different longitudinal designs and different units of analysis. What difficulties can you anticipate with each design?
3. Write a statement for one of your proposed research designs that states how you will ensure adherence to each ethical guideline for the protection of human subjects. Which standards for the protection of human subjects might pose the most difficulty for researchers on your proposed topic? Explain your answers, and suggest appropriate protection procedures for human subjects.

Ethics Questions

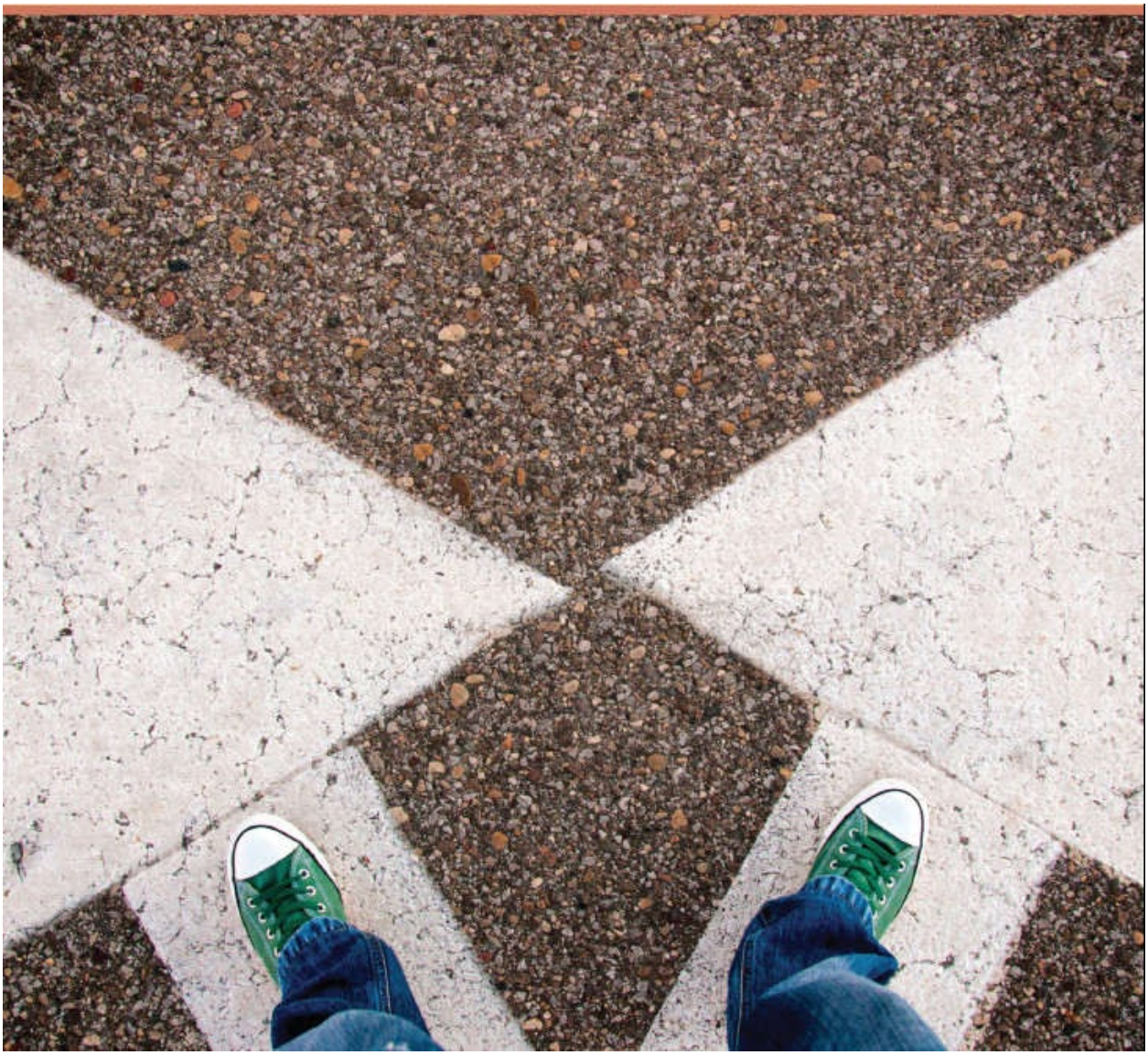
1. Sherman and Berk (1984) and those who replicated their research on the police response to domestic violence assigned persons accused of domestic violence by chance (randomly) to be arrested or not. Their goal was to ensure that the people who were arrested were similar to those who were not arrested. Based on what you now know, do you feel that this random assignment procedure was ethical? Why or why not?
2. Concern with how research results are used is one of the hallmarks of ethical researchers, but deciding what form that concern should take is often difficult. You learned in this chapter about the controversy that occurred after Sherman and Berk (1984) encouraged police departments to adopt a pro-arrest policy in domestic abuse cases based on findings from their Minneapolis study. Do you agree with the researchers' decision, in an effort to minimize domestic abuse, to suggest policy changes to police departments based on their study? Several replication studies failed to confirm the Minneapolis findings. Does this influence your evaluation of what the researchers should have done after the Minneapolis study was completed? What about Sherman's (1992) argument that failure to publicize the Omaha study's finding of the effectiveness of arrest warrants resulted in some cases of abuse that could have been prevented?

Video Interview Questions

Listen to the researcher interview for Chapter 2 at edge.sagepub.com/chamblissmssw5e.

1. What were the research questions that Russ Schutt focused on in the research project about homelessness and housing?
2. Why did they use a randomized experimental design?
3. Schutt stated that the research design was consistent with reasonable ethical standards. Do you agree? Why or why not?
4. What were the answers to the two central research questions, as Schutt described them?
5. To learn more, read Schutt (2011), *Homelessness, Housing, and Mental Illness*, and pay particular attention to the appendix on research methods! http://www.hup.harvard.edu/catalog.php?isbn=9780674051010.

Chapter 3 Ethics in Research



©iStockphoto.com/badmanproduction

Learning Objectives

1. Describe the design of the Milgram obedience experiments and some of the controversies surrounding its methods and results.
2. Identify three other research projects that helped motivate the establishment of human subjects' protections.
3. Define the *Belmont Report*'s three ethical standards for the protection of human subjects.
4. Explain the role of an institutional review board.
5. List current standards for the protection of human subjects in research.
6. Define *debriefing* and review the controversy about the Milgram research.

Imagine this: One spring morning as you are drinking coffee and reading the newspaper, you notice a small ad for a psychology experiment at the local university.

We Will Pay You $45 For One Hour Of Your Time

Persons Needed for a Study of Memory

“Earn money and learn about yourself,” it continues. Feeling a bit bored, you call and schedule an evening visit to the lab.

You are about to enter one of the most ethically controversial experiments in the history of social science.

You arrive at the assigned room at the university and are immediately impressed by the elegance of the building and the professional appearance of the personnel. In the waiting room, you see a man dressed in a lab technician’s coat talking to another visitor, a middle-aged fellow dressed in casual attire. The man in the lab coat turns, introduces himself, and explains that, as a psychologist, he is interested in whether people learn better when they are punished for making mistakes. He quickly convinces you that this is an important question; he then explains that his experiment on punishment and learning will discover the answer. Then he announces, “I’m going to ask one of you to be the teacher here tonight and the other one to be the learner.”

Video Link

Watch excerpts from Milgram’s experiment.

The experimenter (as we’ll refer to him from now on) says he will write either *teacher* or *learner* on small identical slips of paper and then asks both of you to draw one. Yours says *teacher*.

The experimenter now says, in a matter-of-fact way, “All right. Now the first thing we’ll have to do is to set the learner up so that he can get some type of punishment.”

He leads you both behind a curtain, sits the learner in the chair, straps down both of his arms, and attaches an electric wire to his left wrist (Exhibit 3.1). The wire

is connected to a console with 30 switches and a large dial, on the other side of the curtain. When you ask what the wire is for, the experimenter demonstrates. He asks you to hold the end of the wire, walks back to the control console, and flips several switches. You hear a clicking noise, see the dial move, and then feel an electric shock in your hand. When the experimenter flips the next switch, the shock increases.

"Ouch!" you say. "So that's the punishment. Couldn't it cause injury?" The experimenter explains that the machine is calibrated so that it will not cause permanent injury but admits that when turned up all the way, it is very, very painful.

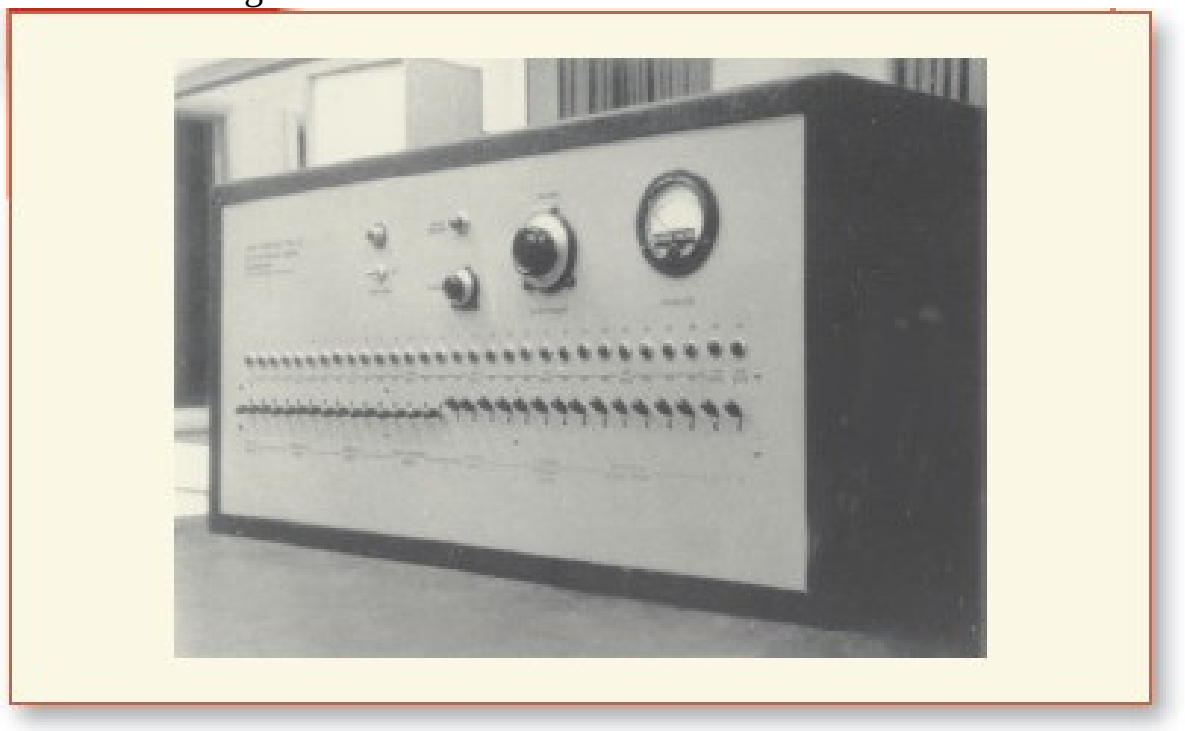
Now you walk back to the other side of the room (so that the learner is behind the curtain) and sit before the console (Exhibit 3.2). The experimental procedure has four simple steps:

1. You read aloud a series of word pairs, such as *blue box, nice day, wild duck*, and so on.
2. You read one of the first words from those pairs and a set of four words, one of which is the original paired word. For example, you might say, "blue: sky-ink-box-lamp."
3. The learner states the word that he thinks was paired with the first word you read (*blue*). If he gives a correct response, you compliment him and move on to the next word. If he makes a mistake, you flip a switch on the console. This causes the learner to feel a shock on his wrist.
4. After each mistake, you are to flip the next switch on the console, progressing from left to right. You note that a label corresponds to every 5th mark on the dial, with the first mark labeled *slight shock*, the 5th mark labeled *moderate shock*, the 10th *strong shock*, and so on through *very strong shock, intense shock, extreme intensity shock*, and *danger: severe shock.*

Exhibit 3.1 Learner Strapped in Chair With Electrodes

Source: From the film Obedience © 1968 by Stanley Milgram, © Renewed 1993 by Alexandra Milgram, and distributed by Penn State Media Sales.

Exhibit 3.2 Milgram's "Shock Generator"



Source: From the film Obedience © 1968 by Stanley Milgram, © Renewed 1993 by Alexandra Milgram, and distributed by Penn State Media Sales

You begin. The learner at first gives some correct answers, but then he makes a few errors. Soon you are beyond the 5th mark (moderate shock) and are moving in the direction of more and more severe shocks. As you turn the dial, the learner's reactions increase in intensity: from a grunt at the 10th mark (strong shock) to painful groans at higher levels, to anguished cries of "get me out of here" at the extreme intensity shock levels, to a deathly silence at the highest level. When you protest at administering the stronger shocks, the experimenter tells you, "The experiment requires that you continue." Occasionally he says, "It is absolutely essential that you continue."

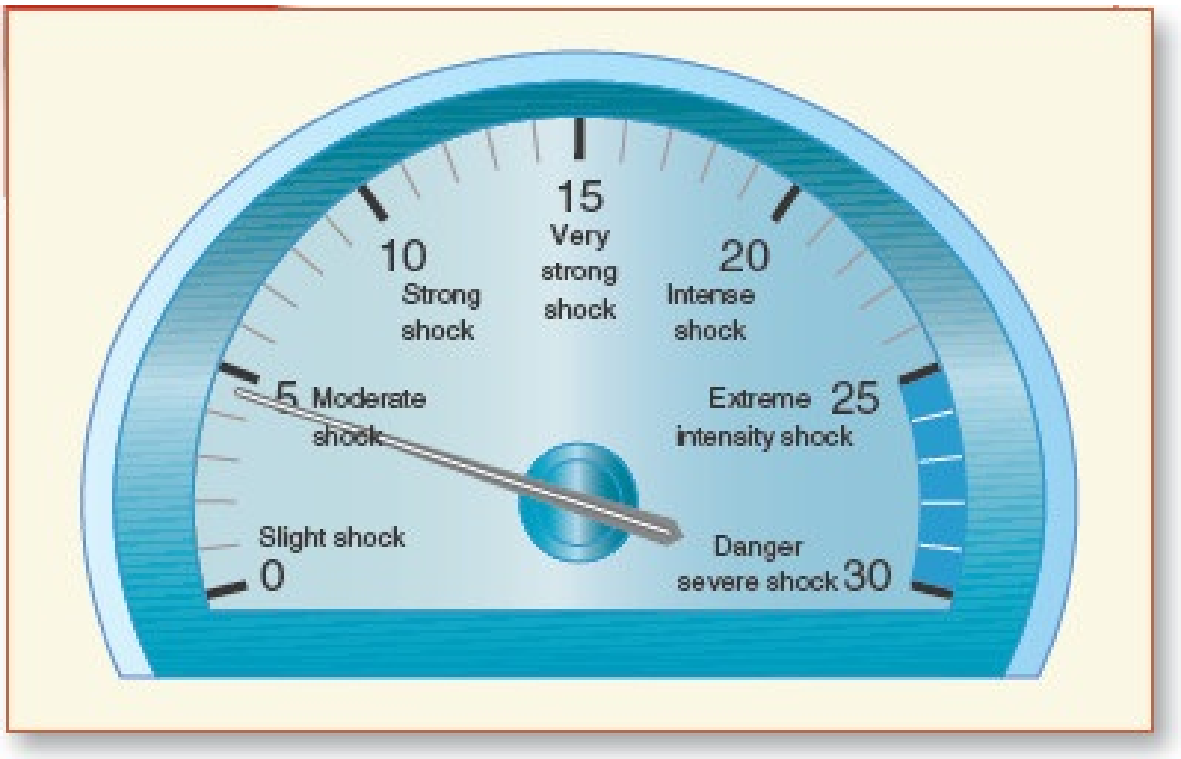
This is a simplified version of the famous Stanley **Milgram's obedience experiments**, begun at Yale University in 1960. Outside the laboratory, Milgram surveyed Yale undergraduates and asked them to indicate at what level they would terminate their "shocks" if they were in the study. Now, please mark on the console below the most severe shock that you would agree to give the learner (Exhibit 3.3).

Obedience experiments (Milgram's): A series of famous experiments conducted during the 1960s by Stanley Milgram, a psychologist from Yale University, testing subjects' willingness to cause pain to another person if instructed to do so.

The average (mean) maximum shock level predicted by the Yale undergraduates was 9.35, corresponding to a strong shock. Only one student predicted that he would provide a stimulus above that level, at the very strong level. Responses were similar from nonstudent groups.

But the actual average level of shock the 40 adults who volunteered for the experiment administered was 24.53—higher than extreme intensity shock and just short of danger: severe shock. Of Milgram's original 40 subjects, 25 complied entirely with the experimenter's demands, going all the way to the top of the scale (labeled simply as *XXX*). Judging from the subjects' visibly high stress, and from their subsequent reports, they believed that the learner was receiving physically painful shocks. **(In fact, no electric shocks were actually delivered.)**

Exhibit 3.3 Shock Meter



We introduce the Milgram experiment not to discuss obedience to authority but instead to introduce research ethics. We refer to Milgram's obedience studies throughout this chapter because they ultimately had as profound an influence on scientists' thinking about ethics as on how we understand obedience to authority. Although Milgram died in 1984, the controversy around his work did not. A recent review of the transcripts and interviews with many participants raises additional concerns even about the experiment's scientific validity, as well as its ethics (Perry 2013).

Throughout this book, we discuss ethical problems common to various research methods; in this particular chapter, we present in more detail some of the general ethical principles that professional social scientists use in monitoring their work.

Nuremberg war crime trials: Trials held in Nuremberg, Germany, in the years following World War II, in which the former leaders of Nazi Germany were charged with war crimes and crimes against humanity; frequently considered the first trials for people accused of genocide.

Tuskegee syphilis study: Research study conducted by a branch of the U.S. government, lasting for roughly 50 years (ending in the 1970s), in which a sample of African American men diagnosed with syphilis were deliberately left untreated, without their knowledge, to learn about the lifetime course of the disease.

Historical Background

Formal procedures for the protection of participants in research grew out of some widely publicized abuses. A defining event occurred in 1946, when the **Nuremberg war crime trials** exposed horrific medical experiments conducted during World War II by Nazi doctors in the name of "science." During the 1950s and 1960s, American military personnel and Pacific Islanders were sometimes unknowingly exposed to radiation during atomic bomb tests. And in the 1970s, Americans were shocked to learn that researchers funded by the U.S. Public Health Service had, for decades, studied 399 low-income African American men diagnosed with syphilis in the 1930s to follow the "natural" course of the illness (Exhibit 3.4). In the **Tuskegee syphilis study**, many participants were not informed of their illness and were denied treatment until 1972, even though a cure (penicillin) was developed in the 1950s (Jones 1993).

Audio Link

Listen to a case of unethical human treatment.

Such egregious violations of human rights resulted, in the United States, in the creation of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The commission's 1979 ***Belmont Report*** (U.S. Department of Health, Education, and Welfare 1979) established three basic ethical principles for the protection of human subjects (Exhibit 3.5):

1. **Respect for persons**—treating persons as autonomous agents and protecting those with diminished autonomy
2. **Beneficence**—minimizing possible harms and maximizing benefits
3. **Justice**—distributing benefits and risks of research fairly

Exhibit 3.4 Tuskegee Syphilis Experiment

Source: Tuskegee Syphilis Study Administrative Records. Records of the Centers for Disease Control and Prevention. National Archives—Southeast Region (Atlanta).

The Department of Health and Human Services and the Food and Drug Administration then translated these principles into specific regulations, which were adopted in 1991 as the **Federal Policy for the Protection of Human Subjects**. This policy has shaped the course of social science research ever since, and you will have to consider it as you design your own research investigations. Some professional associations—such as the American Psychological Association, the American Political Science Association, the American Sociological Association, university review boards, and ethics committees in other organizations—set standards for the treatment of human subjects by their members, employees, and students; these standards are designed to comply with the federal policy.

Federal regulations require that every institution that seeks federal funding for biomedical or behavioral research on human subjects have an **institutional**

review board (IRB) that reviews research proposals. If you do research for a class assignment, you may need to prepare a brief IRB proposal, so board members can be sure that your project meets all ethical standards. IRBs at universities and other agencies apply ethics standards that are set by federal regulations but can be expanded or made more specific by the IRB itself (Sieber 1992: 5, 10). To promote adequate review of ethical issues, the regulations require that IRBs include members with diverse backgrounds. The **Office for Protection From Research Risks** in the **National Institutes of Health** monitors IRBs, with the exception of research involving drugs (which is the responsibility of the federal Food and Drug Administration).

Belmont Report: Report in 1979 of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research stipulating three basic ethical principles for the protection of human subjects: respect for persons, beneficence, and justice.

Respect for persons: In human subjects ethics discussions, treating persons as autonomous agents and protecting those with diminished autonomy.

Beneficence: Minimizing possible harms and maximizing benefits.

Justice: As used in human research ethics discussions, distributing benefits and risks of research fairly.

Federal Policy for the Protection of Human Subjects: Federal regulations codifying basic principles for conducting research on human subjects; used as the basis for professional organizations' guidelines.

Institutional review board (IRB): A group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to subjects.

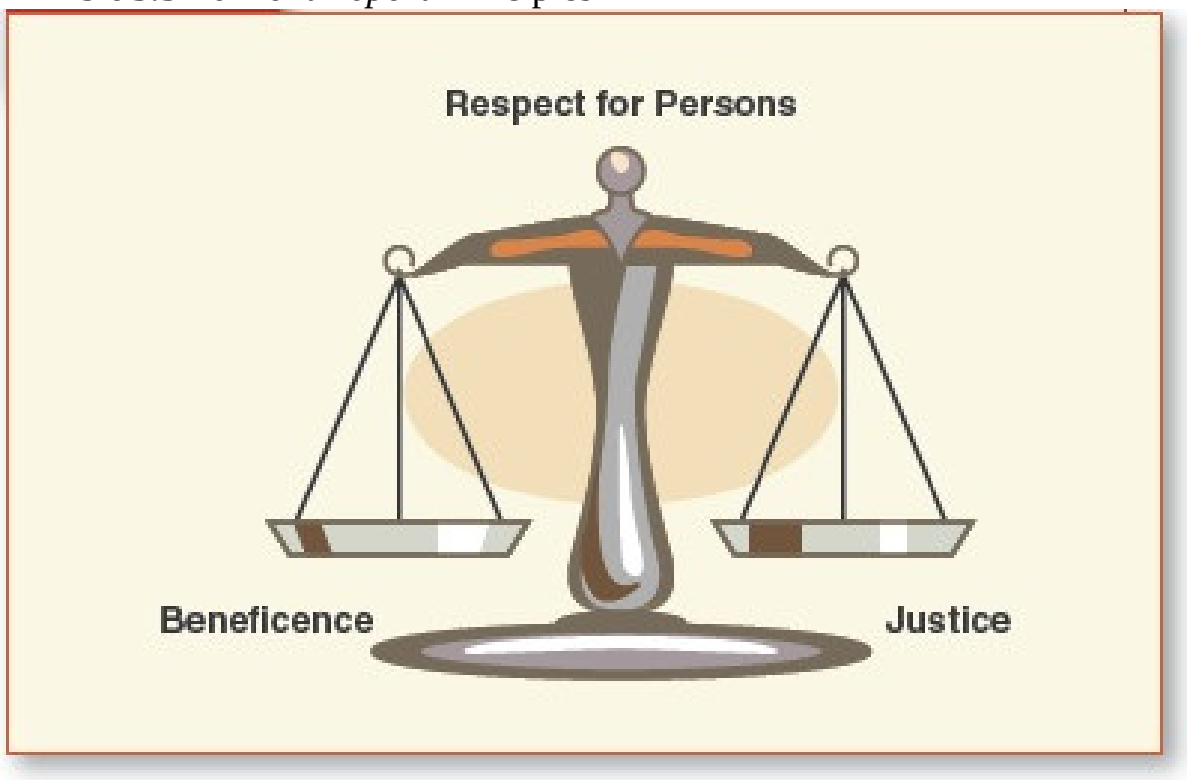
Office for Protection From Research Risks, National Institutes of Health: Federal agency that monitors institutional review boards (IRBs).

Ethical Principles

The American Sociological Association (ASA), like other professional social science organizations, has adopted, for practicing sociologists, ethical guidelines that are more specific than the federal regulations. Professional organizations may also review complaints of unethical practices when asked.

The *Code of Ethics* of the ASA (1997) is summarized at the ASA website (www.asanet.org); the complete text of the code is also available at this site.

Exhibit 3.5 *Belmont Report* Principles



Source: U.S. Department of Health, Education, and Welfare 1979.

Mostly, ethical issues in research are covered by four guidelines:

1. To protect research subjects
2. To maintain honesty and openness
3. To achieve valid results

4. To encourage appropriate application

Each of these guidelines became a focus of the debate about Milgram's experiments, to which we will refer frequently. Did Milgram respect the spirit expressed in these principles? You will find that there is no simple answer to the question of what is (or isn't) ethical research practice.

Protecting Research Subjects

This guideline, our most important, can be divided into four specific directions:

1. Avoid harming research participants.
2. Obtain informed consent.
3. Avoid deception in research, except in limited circumstances.
4. Maintain privacy and confidentiality.

Avoid Harming Research Participants

This standard may seem straightforward, but can be difficult to interpret in specific cases. Does it mean that subjects should not be harmed even mentally or emotionally? That they should feel no anxiety or distress?

The most serious charge leveled against the ethics of Milgram's study was that he had harmed his subjects. A verbatim transcript of one session will give you an idea of what participants experienced as the "shock generator," which made it appear they were delivering increasingly severe shocks to the learner (Milgram 1965: 67):

> *150 volts delivered.* You want me to keep going?
> *165 volts delivered.* That guy is hollering in there. . . . He's liable to have a heart condition. You want me to go on?
> *180 volts delivered.* He can't stand it! I'm not going to kill that man in there! You hear him hollering? He's hollering. He can't stand it. . . . I mean who is going to take responsibility if anything happens to that gentleman? *[The experimenter accepts responsibility.]* All right.
> *195 volts delivered.* You see he's hollering. Hear that. Gee, I don't know. *[The experimenter says: "The experiment requires that you go on."]* I know it does, sir, but I mean—phew—he don't know what he's in for. He's up to 195 volts.
> *210 volts delivered.*
> *225 volts delivered.*
> *240 volts delivered.*

The experimental manipulation generated "extraordinary tension" (Milgram

1963: 377):

> Subjects were observed to sweat, tremble, stutter, bite their lips, groan and dig their fingernails into their flesh. . . . Full-blown, uncontrollable seizures were observed for 3 subjects. One . . . seizure so violently convulsive that it was necessary to call a halt to the experiment [for that individual]. (p. 375)

An observer (behind a one-way mirror) reported, "I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse" (Milgram 1963: 377).

Milgram's "Behavioral Study of Obedience" was published in 1963 in the *Journal of Abnormal and Social Psychology*. The next year, the *American Psychologist* published a critique of the experiment's ethics by psychologist Diana Baumrind (1964: 421). From Baumrind's perspective, the emotional disturbance in subjects was "potentially harmful because it could easily effect an alteration in the subject's self-image or ability to trust adult authorities in the future" (p. 422). Milgram (1964) quickly countered,

> Momentary excitement is not the same as harm. As the experiment progressed there was no indication of injurious effects in the subjects; and as the subjects themselves strongly endorsed the experiment, the judgment I made was to continue the experiment. (p. 849)

Milgram (1963) also attempted to minimize harm to subjects with postexperiment procedures "to assure that the subject would leave the laboratory in a state of well being" (p. 374). A friendly reconciliation was arranged between the subject and the victim, and an effort was made to reduce any tensions that arose as a result of the experiment.

In some cases, the "dehoaxing"—or **debriefing**—discussion was extensive, and all subjects were promised (and later received) a comprehensive report (Milgram 1964: 849). But Baumrind (1964) was unconvinced: "It would be interesting to know what sort of procedures could dissipate the type of emotional disturbance just described" (p. 422).

When Milgram (1964: 849) surveyed subjects in a follow-up, 83.7% endorsed the statement that they were “very glad” or “glad” “to have been in the experiment,” 15.1% were “neither sorry nor glad,” and just 1.3% were “sorry” or “very sorry” to have participated. Interviews by a psychiatrist a year later found no evidence “of any traumatic reactions” (Milgram 1974: 197). Subsequently, Milgram argued, “The central moral justification for allowing my experiment is that it was judged acceptable by those who took part in it” (Milgram as cited in Cave & Holm 2003: 32).

Video Link

Watch excerpts from Zimbardo’s Stanford prison experiment.

In a later article, Baumrind (1985: 168) dismissed the value of the self-reported “lack of harm” of subjects who had been willing to participate in the experiment and noted that 16% did *not* endorse the statement that they were “glad” they had participated in the experiment. Many social scientists, ethicists, and others concluded that Milgram’s procedures had not harmed subjects and so were justified by the knowledge they produced; others sided with Baumrind’s criticisms (Miller 1986: 88–138).

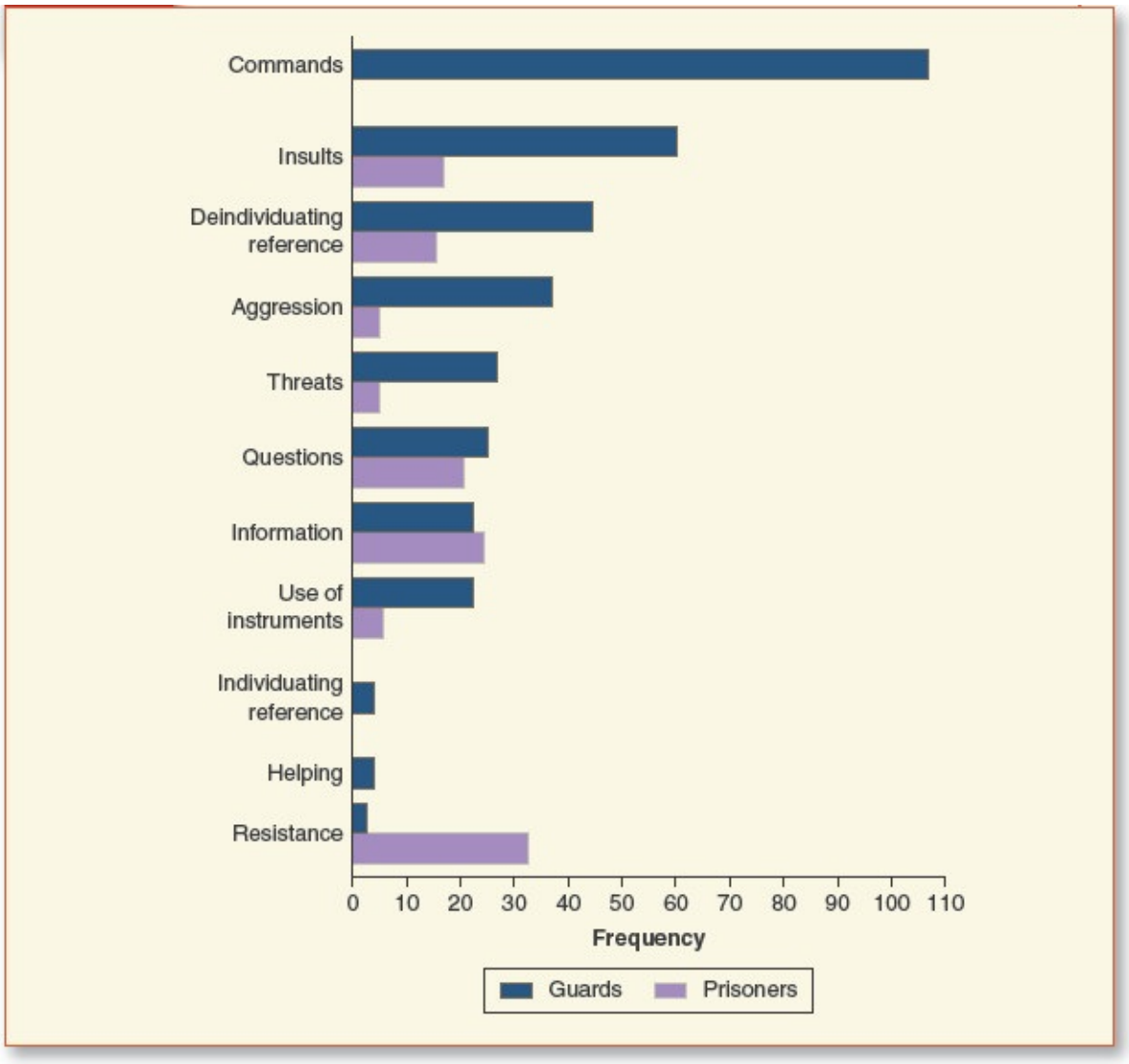
Or, consider the possible harm to subjects in the famous prison simulation study at Stanford University (Haney, Banks, & Zimbardo 1973). Philip **Zimbardo’s prison simulation study** was designed to investigate the impact of being either a guard or a prisoner in a prison, a “total institution.” The researchers selected apparently stable and mature young male volunteers and asked them to sign a contract to work for 2 weeks as a guard or a prisoner in a simulated prison. Within the first 2 days after the prisoners were incarcerated in a makeshift basement prison, the prisoners began to be passive and disorganized, and the guards became “sadistic”—verbally and physically aggressive (Exhibit 3.6). Five “prisoners” were soon released for depression, uncontrollable crying, fits of rage, and, in one case, a psychosomatic rash. Instead of letting things continue for 2 weeks as planned, Zimbardo and his colleagues terminated the experiment after 6 days to avoid harming subjects.

Participants playing the prisoner role certainly felt some stress, but postexperiment discussion sessions seemed to relieve this; follow-up during the next year indicated no lasting negative effects on the participants and some

benefits in the form of greater insight. And besides, Zimbardo and his colleagues had no way of predicting the bad outcome; indeed, they were themselves surprised (Haney et al. 1973).

Withholding beneficial treatment can be another way of causing harm to subjects. Sometimes, in an ethically debatable practice, researchers will actually withhold treatments from some subjects, knowing that those treatments would probably help the people, to accurately measure *how much* they helped. For example, in some recent studies of AIDS drugs conducted in Africa, researchers provided different levels of AIDS-combating drugs to different groups of patients with the disease. Some patients received no drug therapy at all, even though all indications were that the drug treatments would help them. From the point of view of pure science, this makes sense: You can't really know how effective the drugs are unless you try different treatments on different people who start from the same situation (e.g., having AIDS). But the research has provoked a tremendous outcry across the world because many people find the practice of deliberately not treating people—in particular, impoverished black people living in Third World countries—to be morally repugnant.

Exhibit 3.6 Chart of Guard and Prisoner Behavior



Source: From The Lucifer Effect by Philip G. Zimbardo. Copyright 2007 by Philip G. Zimbardo, Inc. Used by permission of Random House, an imprint and division of Random House LLC, and Random House Group Ltd. All rights reserved.

Audio Link

Listen to how Zimbardo's research continues to inform ethics.

Even well-intentioned researchers may fail to foresee potential ethical problems. Milgram (1974: 27–31) reported that he and his colleagues were surprised by the subjects' willingness to administer such severe shocks. In Zimbardo's prison simulation, all the participants signed consent forms, but even the researchers did not realize that participants would fall apart so quickly, that some prisoners would have to be released within a few days, or that others would soon be begging to be released from the mock prison. Some risks cannot be foreseen, so they cannot be consented to.

Journal Link

Read about the written consent needed for a youth smoking prevention trial.

Debriefing: A researcher's informing subjects after an experiment about the experiment's purposes and methods and evaluating subjects' personal reactions to the experiment.

Prison simulation study (Zimbardo's): Famous study from the early 1970s, organized by Stanford psychologist Philip Zimbardo, demonstrating the willingness of average college students quickly to become harsh disciplinarians when put in the role of (simulated) prison guards over other students; usually interpreted as demonstrating an easy human readiness to become cruel.

Obtain Informed Consent

Just defining *informed consent* may also be more difficult than it first appears. To be informed, consent must be given by persons who are competent to consent, have consented voluntarily, are fully informed about the research, and have comprehended what they have been told (Reynolds 1979). Yet, you probably realize, as did Baumrind (1985), that because of the inability to communicate perfectly, "Full disclosure of everything that could possibly affect a given subject's decision to participate is not possible, and therefore cannot be ethically required" (p. 165).

Obtaining informed consent creates additional challenges for researchers. For instance, the language of the consent form must be clear and understandable yet sufficiently long and detailed to explain what will actually happen in the research. Examples A (Exhibit 3.7) and B (Exhibit 3.8) illustrate two different approaches to these trade-offs. Consent form A was approved by a university for

a substance abuse survey with undergraduate students. It is brief and to the point but leaves quite a bit to the imagination of the prospective participants. Consent form B reflects the requirements of an academic hospital's IRB. Because the hospital is used to reviewing research proposals involving drugs and other treatment interventions with hospital patients, it requires a very detailed and lengthy explanation of procedures and related issues, even for a simple survey. Requiring prospective participants to sign such lengthy forms can reduce their willingness to participate in research and perhaps influence their responses if they do agree to participate (Larson 1993: 114).

Research|Social Impact Link

Read about deception and control in research.

When an experimental design requires subject deception, researchers may withhold information before the experiment but then debrief subjects after the experiment ends (Milgram did this). In the debriefing, the researcher explains what really happened in the experiment, and why, and responds to subjects' questions. A carefully designed debriefing procedure can often help research participants deal with their anger or embarrassment at having been deceived (Sieber 1992: 39–41), thus substituting for fully informed consent before the experiment.

Interactive Exercises

Ethical Issues

Exhibit 3.7 Consent Form A

University of Massachusetts Boston
Department of Sociology
October 28, 2014

Dear _____:

The health of students and their use of alcohol and drugs are important concerns for every college and university. The enclosed survey is about these issues at UMass/Boston. It is sponsored by University Health Services and the PRIDE Program (Prevention, Resources, Information, and Drug Education). The questionnaire was developed by graduate students in Applied Sociology, Nursing, and Gerontology.

You were selected for the survey with a scientific, random procedure. Now it is important that you return the questionnaire so that we can obtain an unbiased description of the undergraduate student body. Health Services can then use the results to guide campus education and prevention programs.

The survey requires only about 20 minutes to complete. Participation is completely voluntary and anonymous. No one will be able to link your survey responses to you. In any case, your standing at the University will not be affected whether or not you choose to participate. Just be sure to return the enclosed postcard after you mail the questionnaire so that we know we do not have to contact you again.

Please return the survey by November 15th. If you have any questions or comments, call the PRIDE program at 287–5680 or Professor Schutt at 287–6250. Also call the PRIDE program if you would like a summary of our final report.

Thank you in advance for your assistance.

Russell K. Schutt, PhD
Professor and Chair

Exhibit 3.8 Consent Form B

Protocol Title: ASSESSING COMMUNITY HEALTH WORKERS' ATTITUDES AND KNOWLEDGE ABOUT EDUCATING COMMUNITIES ABOUT CANCER CLINICAL TRIALS

DF/HCC Principal Research Investigator / Institution: Dr. Russell Schutt, PhD/Beth Israel Deaconess Medical Center and Univ. of Massachusetts, Boston

DF/HCC Site-Responsible Research Investigator(s) / Institution(s): Lidia Schapira, MD/Massachusetts General Hospital

Interview Consent Form

A. INTRODUCTION

We are inviting you to take part in a research study. Research is a way of gaining new knowledge. A person who participates in a research study is called a "subject." This research study is evaluating whether community health workers might be willing and able to educate communities about the pros and cons of participating in research studies.

It is expected that about 10 people will take part in this research study.

An institution that is supporting a research study either by giving money or supplying something that is important for the research is called the "sponsor." The sponsor of this protocol is National Cancer Institute and is providing money for the research study.

This research consent form explains why this research study is being done, what is involved in participating in the research study, the possible risks and benefits of the research study, alternatives to participation, and your rights as a research subject. The decision to participate is yours. If you decide to participate, please sign and date at the end of the form. We will give you a copy so that you can refer to it while you are involved in this research study.

If you decide to participate in this research study, certain questions will be asked of you to see if you are eligible to be in the research study. The research study has certain requirements that must be met. If the questions show that you can be in the research study, you will be able to answer the interview questions.

If the questions show that you cannot be in the research study, you will not be able to participate in this research study.

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

We encourage you to take some time to think this over and to discuss it with other people and to ask questions now and at any time in the future.

B. WHY IS THIS RESEARCH STUDY BEING DONE?

Deaths from cancer in general and for some specific cancers are higher for black people compared to white people, for poor persons compared to nonpoor persons, and for rural residents compared to non-rural residents. There are many reasons for higher death rates between different subpopulations. One important area for changing this is to have more persons from minority groups participate in research about cancer. The process of enrolling minority populations into clinical trials is difficult and does not generally address the needs of their communities. One potential way to increase particpation in research is to use community health workers to help educate communities about research and about how to make sure that researchers are ethical. We want to know whether community health workers think this is a good strategy and how to best carry it out.

C. WHAT OTHER OPTIONS ARE THERE?

Taking part in this research study is voluntary. Instead of being in this research study, you have the following option:

- Decide not to participate in this research study.

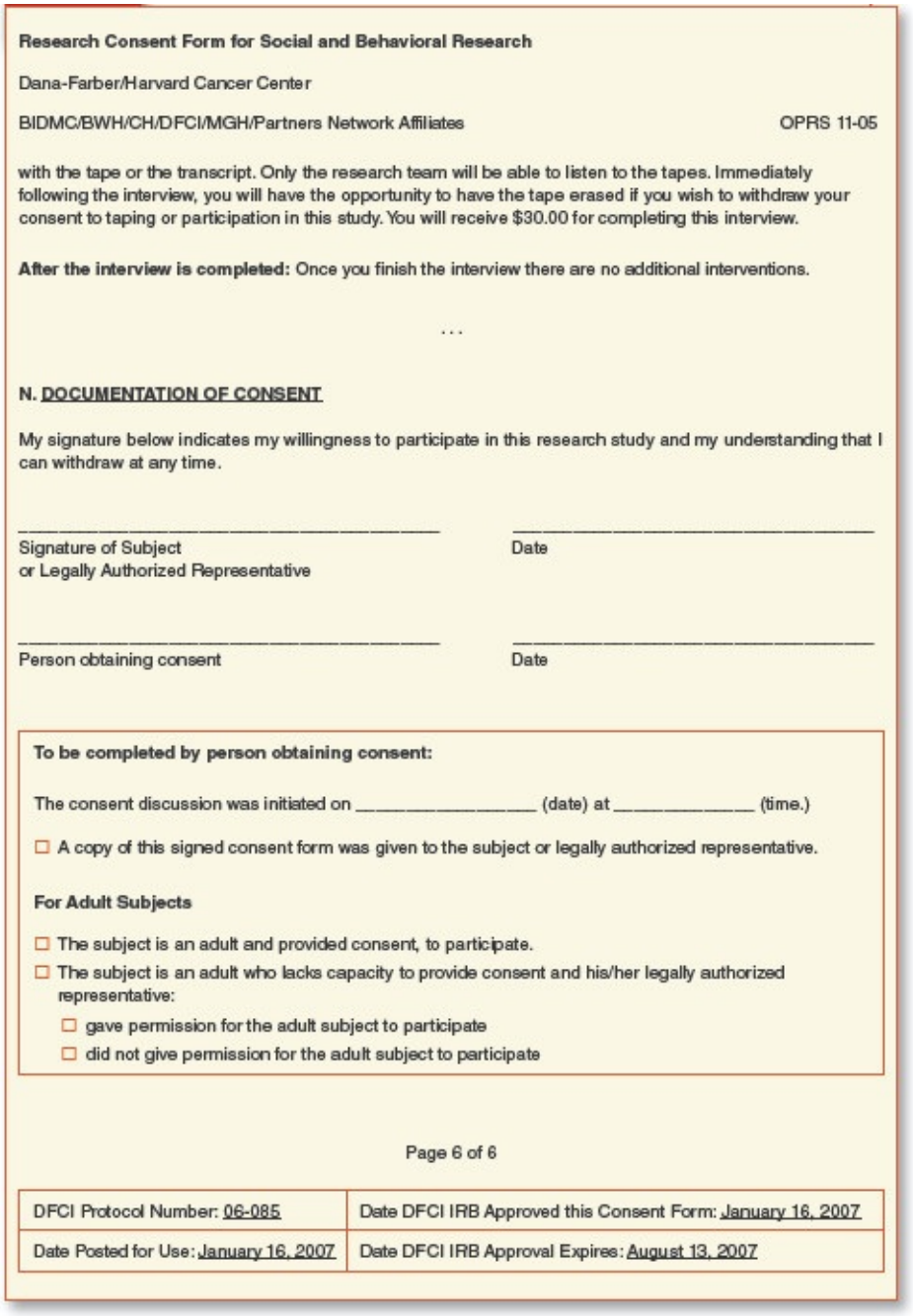
D. WHAT IS INVOLVED IN THE RESEARCH STUDY?

Before the research starts (screening): After signing this consent form, you will be asked to answer some questions about where you work and the type of community health work you do to find out if you can be in the research study.

If the answers show that you are eligible to participate in the research study, you will be eligible to participate in the research study. If you do not meet the eligibility criteria, you will not be able to participate in this research study.

After the screening procedures confirm that you are eligible to participate in the research study: You will participate in an interview by answering questions from a questionnaire. The interview will take about 90 minutes. If there are questions you prefer not to answer we can skip those questions. The questions are about the type of work you do and your opinions about participating in research. If you agree, the interview will be taped and then transcribed. Your name and no other information about you will be associated:

. . .



Research Consent Form for Social and Behavioral Research

Dana-Farber/Harvard Cancer Center

BIDMC/BWH/CH/DFCI/MGH/Partners Network Affiliates OPRS 11-05

with the tape or the transcript. Only the research team will be able to listen to the tapes. Immediately following the interview, you will have the opportunity to have the tape erased if you wish to withdraw your consent to taping or participation in this study. You will receive $30.00 for completing this interview.

After the interview is completed: Once you finish the interview there are no additional interventions.

. . .

N. DOCUMENTATION OF CONSENT

My signature below indicates my willingness to participate in this research study and my understanding that I can withdraw at any time.

_____ _____

Signature of Subject or Legally Authorized Representative — Date

_____ _____

Person obtaining consent — Date

To be completed by person obtaining consent:

The consent discussion was initiated on _____ (date) at _____ (time.)

☐ A copy of this signed consent form was given to the subject or legally authorized representative.

For Adult Subjects

☐ The subject is an adult and provided consent, to participate.

☐ The subject is an adult who lacks capacity to provide consent and his/her legally authorized representative:

- ☐ gave permission for the adult subject to participate
- ☐ did not give permission for the adult subject to participate

Page 6 of 6

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

Finally, some participants can't truly give informed consent. College students, for instance, may feel unable to refuse if their professor asks them to be in an experiment. Legally speaking, children cannot give consent to participate in research; a child's legal guardian must give written informed consent to have the child participate in research (Sieber 1992). Then, the child must in most circumstances be given the opportunity to give or withhold *assent* to participate in research, usually by a verbal response to an explanation of the research. Special protections exist for other vulnerable populations—prisoners, pregnant

women, mentally disabled persons, and educationally or economically disadvantaged persons. And in a sense, anyone deliberately deceived in an experiment cannot be said to really have given "informed" consent, since the person wasn't honestly told what would happen.

Social media and digital technologies have in recent years opened the doors to new kinds of ethical problems in research, by blurring the lines between public and private behavior. If you have a Facebook or Myspace page with 600 "friends," is that your private page, or a public document? In Chapter 8, we'll see how social researchers are eagerly mining such data for information on people's social networks; "Employers are looking at people's online postings and Googling information about them, and I think researchers are right behind them," said Professor Nicholas Christakis (as cited in Rosenbloom 2007: 2), a Harvard sociologist in a *New York Times* article in 2007. But the federal guidelines under which institutional review boards are set up didn't anticipate the Internet. "The [human subject] rules were made for a different world, a pre-Facebook world," said Samuel D. Gosling, a psychology professor at the University of Texas who uses Facebook as a data source. "There is a rule that you are allowed to observe public behavior, but it's not clear if online behavior is public or not" (as cited in Rosenbloom 2007: 2).

Research|Social Impact Link

Read more about issues in attaining and giving informed consent.

In truth, though, the public versus private debate is a long-standing issue in social science. Laud Humphreys (1970) decided that truly informed consent would be impossible to obtain for his study of the social background of men who engage in homosexual behavior in public facilities. Humphreys served as a lookout—a "watch queen"—for men who were entering a public bathroom in a city park with the intention of having sex. In a number of cases, he then left the bathroom and copied the license plate numbers of the cars driven by the men. One year later, he visited the homes of the men and interviewed them as part of a larger study of social issues. Humphreys changed his appearance so that the men did not recognize him. In his book ***Tearoom Trade***, Humphreys concluded that the men who engaged in what were widely viewed as deviant acts were, for the most part, married, suburban men whose families were unaware of their sexual practices. But debate has continued ever since about Humphreys's failure to tell

the men what he was really doing in the bathroom or why he had come to their homes for the interview. He was criticized by many, including some faculty members at the University of Washington who urged that his doctoral degree be withheld. However, many other professors and some members of the gay community praised Humphreys for helping normalize conceptions of homosexuality (Miller 1986: 135).

If you served on your university's IRB, would you allow research such as Humphreys's to be conducted?

***Tearoom Trade*:** Book by Laud Humphreys investigating the social background of men who engage in homosexual behavior in public facilities; controversially, he did not obtain informed consent from his subjects.

Avoid Deception in Research, Except in Limited Circumstances

Deception occurs when subjects are misled about research procedures. Frequently, this is done to simulate real-world conditions in the lab. The goal is to get subjects "to accept as true what is false or to give a false impression" (Korn 1997: 4). In Milgram's (1964) experiment, for example, deception seemed necessary because actually giving electric shocks to the "stooge" would be cruel. Yet, to test obedience, the task had to be troubling for the subjects. Milgram (1974: 187–188) insisted that the deception was absolutely essential. Many other psychological and social psychological experiments would be worthless if subjects understood what was really happening to them while the experiment was in progress. But is this sufficient justification to allow the use of deception?

Some important topics have been cleverly studied using deception. Gary Marshall and Philip Zimbardo (of prison study fame), in a 1979 study, told the student volunteers that they were being injected with a vitamin supplement to test its effect on visual acuity (Korn 1997: 2–3). But to determine the physiological basis of emotion, they actually injected them with adrenaline, so that their heart rate and sweating would increase, and then placed them in a room with a student stooge who acted silly. Jane Allyn Piliavin and Irving Piliavin, in a 1972 study, staged fake seizures on subway trains to study helpfulness (Korn 1997: 3–4). Again, would you allow such deceptive practices if you were a member of your university's IRB? Giving people stimulating drugs, apart from

the physical dangers, is using their very bodies for research without their knowledge. Faking an emergency may lessen one's willingness to help in the future or may, in effect, punish the research subjects—through embarrassment—for their reaction to what is really "just an experiment."

Researcher Interview Link

Watch a researcher describe how an IRB works.

But perhaps risk, not deception per se, is the real problem. Elliot Aronson and Judson Mills's (1959) study of severity of initiation to groups is a good example of experimental research that does not pose greater-than-everyday risks to subjects but still uses deception. This study was conducted at an all-women's college in the 1950s. The student volunteers who were randomly assigned to the "severe initiation" experimental condition had to read a list of embarrassing words. Even in the 1950s, reading a list of potentially embarrassing words in a laboratory setting, then listening to a taped discussion, was unlikely to increase the risks to which students were exposed in their everyday lives. Moreover, the researchers informed subjects that they would be expected to talk about sex and could decline to participate in the experiment if this requirement would bother them. None dropped out. To further ensure that no psychological harm was caused, Aronson and Mills explained the true nature of the experiment to subjects after the experiment. The subjects did not seem perturbed: "None of the Ss expressed any resentment or annoyance at having been misled. In fact, the majority were intrigued by the experiment, and several returned at the end of the academic quarter to ascertain the result" (p. 179).

Are you satisfied that this procedure caused no harm? The minimal deception in the Aronson and Mills experiment, coupled with the lack of any ascertainable risk to subjects and a debriefing, satisfies the ethical standards for research of most psychologists and IRBs, even today.

Journal Link

Read more about how qualitative researchers navigate ensuring privacy among their subjects.

Maintain Privacy and Confidentiality

Maintaining privacy and confidentiality after a study is completed is another way to protect subjects, and the researcher's commitment to that standard should be included in the informed consent agreement (Sieber 1992). Procedures to protect each subject's privacy, such as locking records and creating special identifying codes, must be created to minimize the risk of access by unauthorized persons. For the protection of health care data, the **Health Insurance Portability and Accountability Act (HIPAA)**, passed by Congress in 1996, created much more stringent regulations. As implemented by the U.S. Department of Health and Human Services in 2000 (and revised in 2002), the HIPAA Final Privacy Rule applies to oral, written, and electronic information that "relates to the past, present, or future physical or mental health or condition of an individual" (Legal Information Institute, 2006. § 1320d[6][B]). The HIPAA Rule requires that researchers have valid authorization for any use or disclosure of "protected health information" (PHI) from a health care provider. Waivers of authorization can be granted in special circumstances (Cava, Cushman, & Goodman 2007).

However, statements about confidentiality should be realistic. In 1993, sociologist Rik Scarce was jailed for 5 months for contempt of court after refusing to testify to a grand jury about so-called ecoterrorists. Scarce, a PhD candidate at Washington State University at the time, was researching radical environmentalists and may have had information about a 1991 "liberation" raid on an animal research lab at Washington State. Scarce was eventually released from jail, but he never did violate the **confidentiality** he claimed to have promised his informants (Scarce 2005). Laws allow research records to be subpoenaed and may require reporting child abuse. A researcher also may feel compelled to release information if a health- or life-threatening situation arises and participants need to be alerted.

The National Institutes of Health can issue a **Certificate of Confidentiality** to protect researchers from being legally required to disclose confidential information. Researchers who focus on high-risk populations or behaviors or sensitive topics, such as crime, substance abuse, sexual activity, or genetic information, can request such a certificate. Suspicions of child abuse or neglect must still be reported, and in some states, researchers may still be required to report such crimes as elder abuse (Arwood & Panicker 2007).

Health Insurance Portability and Accountability Act (HIPAA): A U.S. federal law passed in 1996 that guarantees, among other things, specified privacy rights for medical patients, in particular those in research settings.

Confidentiality: Provided by research in which identifying information that could be used to link respondents to their responses is available only to designated research personnel for specific research needs.

Certificate of Confidentiality: Document issued by the National Institutes of Health to protect researchers from being legally required to disclose confidential information.

Research That Matters

You are driving on the highway at about 3 p.m. on a Friday when you see a police officer standing by his squad car, lights flashing. The officer motions you to pull off the road and stop in an area marked off with traffic cones. You are both relieved and surprised when someone in plain clothes working with the police officer then walks over to your car and asks if you would consent to be in a survey. You then notice two large signs that say NATIONAL ROADSIDE SURVEY and VOLUNTARY SURVEY. You are offered $10 to provide an oral fluid sample and answer a few additional questions on drug use.

This is what happened to 10,909 U.S. motorists between July 20 and December 1, 2007, at sites across the United States. Those who agreed to the oral fluid collection were also offered an additional $5 to complete a short alcohol and drug-use disorder questionnaire. Before they drove off, participants were also offered a $50 incentive for providing a blood sample. Drivers who were found to be too impaired to be able to drive safely (blood alcohol level above .05) were given a range of options, including switching with an unimpaired passenger, getting a free ride home, or spending a night in a local motel (at no expense to them). None were arrested or given citations, and no crashes occurred in relation to the study. Those younger than 21 years and those who were pregnant were given informational brochures because of the special risk they face if they consume alcohol.

John H. Lacey and others from the Pacific Institute for Research and Evaluation, C. Debra Furr-Holden from Johns Hopkins University, and Amy Berning from the National Highway Traffic Safety Administration (NHTSA, which funded the study) reported the procedures for this survey in a 2011 article in the *Evaluation Review.* The survey explained that all data collected were maintained as anonymous, so no research participants could be linked to their survey

The 2007 National Roadside Survey identified 10.5% of the drivers as using illegal drugs and 3% as having taken medications.

Source: Lacey, John H., Tara Kelley-Baker, Robert B. Voas, Eduardo Romano, C. Debra Furr-Holden, Pedro Torres, and Amy Berning. 2011. Alcohol- and drug-involved driving in the United States: Methodology for the 2007 National Roadside Survey. *Evaluation Review* 35: 319–353.

Maintaining Honesty and Openness

Protecting subjects, then, is the primary focus of research ethics. But researchers have obligations to other groups, including the scientific community, whose concern with validity requires that scientists be open in disclosing their methods and honest in presenting their findings. To assess the validity of a researcher's conclusions and the ethics of this researcher's procedures, you need to know how the research was conducted. This means that articles or other reports must include a detailed methodology section, perhaps supplemented by appendixes containing the research instruments or websites or other contact information where more information can be obtained. Biases or political motives should be acknowledged because research distorted by political or personal pressures to find particular outcomes is unlikely to be carried out in an honest and open fashion.

Gina Perry's (2013) *Behind the Shock Machine* challenges Milgram's adherence to the goal of honesty and openness, although his initial 1963 article included a description of study procedures, including details about the procedures involved in the learning task, administration of the "sample shock," the shock instructions and the preliminary practice run, the standardized feedback from the "victim" and from the experimenter, and the measures used. Many more details, including pictures, were provided in Milgram's (1974) subsequent book. Perry, though, has revealed misleading statements in Milgram's reports.

The act of publication itself is a vital element in maintaining openness and honesty, since then others can review procedures and debate with the researcher. Although Milgram disagreed sharply with Baumrind's criticisms of his experiments, their mutual commitment to public discourse in journals widely available to psychologists resulted in more comprehensive presentation of study procedures and more thoughtful conversation about research ethics. Almost 50 years later, this commentary continues to inform debates about research ethics (Cave & Holm 2003).

And what about the ethics of concealing from your subjects that you're even doing research? Carolyn Ellis (1986) spent several years living in and studying two small fishing communities on Chesapeake Bay in Massachusetts. Living with these "fisher folk," as she called them, she learned quite a few fairly

intimate details about their lives, including their less-than-perfect hygiene habits (many simply smelled bad from not bathing). When the book was published, many townspeople were enraged that Ellis had lived among them and then, in effect, betrayed their innermost secrets without having told them that she was planning to write a book. There was enough detail in the book, in fact, that some of the fisher folk could be identified, and Ellis had never fully disclosed to the fisher folk that she was doing research. The episode stirred quite a debate among professional sociologists as well.

Here's an even more troubling example of hiding one's motives from one's subjects: In the early 1980s, Professor Erich Goode spent three and a half years doing research on the National Association to Aid Fat Americans. Goode was interested primarily in how overweight people managed their identity and enhanced their own self-esteem by forming support groups. Twenty years after the research, in 2002, Goode published an article in which he revealed that in doing the research, he met and engaged in romantic and sexual relationships with more than a dozen women in that organization. There was a heated discussion among the editors and board members of the journal in which the article was published, not only about the ethics of the researcher doing such a thing but also about the ethics of the journal then publishing an article that seemed to take inappropriate advantage of the unusual subject matter.

Despite the need for openness, researchers may hesitate to disclose their procedures or results to prevent others from "stealing" their ideas and taking the credit. However, failure to be open about procedures can result in difficult disputes. In the 1980s, for instance, there was a long legal battle between a U.S. researcher, Robert Gallo, and a French researcher, Luc Montagnier, both of whom claimed credit for discovering the AIDS virus. Eventually the dispute was settled at the highest levels of government, through an agreement announced by U.S. President Ronald Reagan and French Prime Minister Jacques Chirac (Altman 1987). Gallo and Montagnier jointly developed a chronology of discovery as part of the agreement. Enforcing standards of honesty and encouraging openness about research are often the best solutions to such problems.

Research in the News

What Would an IRB Say?

In 2010, New York City Mayor Michael Bloomberg and his health commissioner, Thomas Frieden, unilaterally moved New Yorkers to a lower salt diet. Although numerous research experiments have attempted to find a relationship between low salt diets and improved health, there have been no conclusive results. But New York City administrators required restaurants to impose a cap on salt intake.

1. Is it ethical to base public policies on ambiguous research results? What if the results are definitive? Should these restaurants be using informed consent forms?
2. Could you imagine a real social experiment in New York restaurants to test the value of lowering salt in foods? What steps might an IRB require?

News Source: Adapted from Tierney, John. 2009. Public policy that makes test subjects of us all. *New York Times*, April 7: D1.

Achieving Valid Results

The pursuit of objective knowledge—the goal of validity—justifies our investigations and our claims to the use of human subjects. We have no business asking people to answer questions, submit to observations, or participate in experiments if we are simply trying to trumpet our own prejudices or pursue our personal interests. If, however, we approach our research projects objectively, setting aside our predilections in the service of learning a bit more about human behavior, we can honestly represent our actions as potentially contributing to the advancement of knowledge.

The details in Milgram's 1963 article and 1974 book on the obedience experiments make a compelling case for his commitment to achieving valid results—to learning how obedience influences behavior. In Milgram's (1963) own words,

> It has been reliably established that from 1933–45 millions of innocent persons were systematically slaughtered on command. . . . Obedience is the psychological mechanism that links individual action to political purpose. It is the dispositional cement that binds men to systems of authority. . . . For many persons obedience may be a deeply ingrained behavior tendency. . . . Obedience may [also] be ennobling and educative and refer to acts of charity and kindness, as well as to destruction. (p. 371)

Milgram (1963) then explains how he devised experiments to study the process of obedience in a way that would seem realistic to the subjects and still allow "important variables to be manipulated at several points in the experiment" (p. 372). Every step in the experiment was carefully designed to ensure that subjects received identical stimuli and that their responses were measured carefully.

Milgram's (1963) attention to validity is also apparent in his reflections on "the particular conditions" of his experiment, for, he notes, "Understanding of the phenomenon of obedience must rest on an analysis of [these conditions]" (p. 377). These particular conditions included the setting for the experiment at Yale University, its purported "worthy purpose" to advance knowledge about learning and memory, and the voluntary participation of the subject as well as of the

learner—as far as the subject knew. The importance of some of these "particular conditions" (such as the location at Yale) was then tested in subsequent replications of the basic experiment (Milgram 1965).

However, not all psychologists agreed that Milgram's approach could achieve valid results. Baumrind's (1964) critique begins with a rejection of the external validity—the generalizability—of the experiment. "The laboratory is unfamiliar as a setting and the rules of behavior ambiguous. . . . Therefore, the laboratory is not the place to study degree of obedience or suggestibility, as a function of a particular experimental condition" (p. 423). And so, "the parallel between authority-subordinate relationships in Hitler's Germany and in Milgram's laboratory is unclear" (p. 423).

Milgram (1964) quickly published a rejoinder in which he disagreed with (among other things) the notion that it is inappropriate to study obedience in a laboratory setting: "A subject's obedience is no less problematical because it occurs within a social institution called the psychological experiment" (p. 850).

Milgram (1974: 169–178) also pointed out that his experiment had been replicated in other places and settings with the same results, that there was considerable evidence that subjects had believed that they actually were administering shocks, and that the "essence" of his experimental manipulation—the request that subjects comply with a legitimate authority—was shared with the dilemma faced by people in Nazi Germany and soldiers at the My Lai massacre in Vietnam (Miller 1986: 182–183).

But Baumrind (1985) was still not convinced. In a follow-up article in the *American Psychologist*, she argued that "far from illuminating real life, as he claimed, Milgram in fact appeared to have constructed a set of conditions so internally inconsistent that they could not occur in real life" (p. 171).

Milgram assumed that obedience could fruitfully be studied in the laboratory; Baumrind disagreed. Both, however, buttressed their ethical arguments with assertions about the external validity (or invalidity) of the experimental results. They agreed, in other words, that a research study is partly justified by its valid findings—the knowledge to be gained. If the findings aren't valid, they can't justify the research at all. It is hard to justify any risk for human subjects, or even any expenditure of time and resources, if our findings tell us nothing about human behavior.

Careers and Research

Kristen Kenny, Research Compliance Specialist

Source: Kristen Kenny

Kristen Kenny comes from a long line of musicians and artists and was the first in her family to graduate from college. Kenny majored in filmmaking and performance art at the Massachusetts College of Art and soon started working on small films and in theater doing everything from set design, hair, and makeup to costume design and acting. The arts have their fair share of interesting characters; this was the beginning of Kenny's training in dealing with a variety of difficult personalities and learning how to listen and how to react.

After years of working a variety of jobs in the entertainment field, Kenny found herself working as a receptionist in the music industry, a hotbed of difficult personalities, contracts, and

negotiations. Within a year, Kenny had been promoted to assistant talent buyer for small clubs and festivals in the Boston area. This job helped Kenny develop the skill of reading dense contract documents and being able to identify what contractual clause language stays and what is deleted. Eventually the music industry started to wane and Kenny was laid off, but a friend at a local hospital who was in dire need of someone who could interpret volumes of documents and deal with bold personalities asked her to apply for a job as their IRB administrator. Kenny had no idea what an IRB was, but she attended trainings and conferences to learn the IRB trade. Three years later, Kenny was asked to join the Office of Research and Sponsored Programs at the University of Massachusetts, Boston, as the IRB administrator.

Now, as a research compliance specialist II, Kenny maintains the IRB and other regulatory units and has developed a training curriculum and program for the Office of Research and Sponsored Programs. And if you look hard enough you can find her clothing and fabric designs on eBay, Etsy, and her own website.

Encouraging Appropriate Application

Finally, scientists must consider the uses to which their research is put. Although many scientists believe that personal values should be left outside the laboratory, some feel that it is proper—even necessary—for scientists to concern themselves with the way their research is used.

Milgram made it clear that he was concerned about the phenomenon of obedience precisely because of its implications for people's welfare. As you have already learned, his first article (1963) highlighted the atrocities committed under the Nazis by citizens and soldiers who were "just following orders." In his more comprehensive book on the obedience experiments (1974), he also used his findings to shed light on the atrocities committed in the Vietnam War at My Lai, slavery, the destruction of the American Indian population, and the internment of Japanese Americans during World War II. Milgram makes no explicit attempt to "tell us what to do" about this problem. In fact, as a dispassionate psychological researcher, Milgram (1974) tells us, "What the present study [did was] to give the dilemma [of obedience to authority] contemporary format by treating it as subject matter for experimental inquiry, and with the aim of understanding rather than judging it from a moral standpoint" (p. xi).

Yet it is impossible to ignore the very practical implications of Milgram's investigations. His research highlighted the extent of obedience to authority and identified multiple factors that could be manipulated to lessen blind obedience (such as encouraging dissent by just one group member, removing the subject from direct contact with the authority figure, and increasing the contact between the subject and the victim).

A widely publicized experiment on the police response to domestic violence, mentioned earlier, provides an interesting cautionary tale about the uses of science. Lawrence Sherman and Richard Berk (1984) arranged with the Minneapolis police department for the random assignment of persons accused of domestic violence to be either arrested or simply given a warning. The results of this field experiment indicated that those who were arrested were less likely subsequently to commit violent acts against their partners. Sherman (1993) explicitly cautioned police departments not to adopt mandatory arrest policies based solely on the results of the Minneapolis experiment, but the results were

publicized in the mass media and encouraged many jurisdictions to change their policies (Binder & Meeker 1993; Lempert 1989). Although we now know that the original finding of a deterrent effect of arrest did not hold up in many other cities where the experiment was repeated, Sherman (1992: 150–153) later suggested that implementing mandatory arrest policies might have prevented some subsequent cases of spouse abuse. In particular, in a follow-up study in Omaha, arrest warrants reduced repeat offenses among spouse abusers who had already left the scene when police arrived. However, this Omaha finding was not publicized, so it could not be used to improve police policies. So how much publicity is warranted, and at what point in the research should it occur?

Or what can researchers do if others misinterpret their findings, or use them in misleading ways? For example, during the 1980s, Murray Straus, a prominent researcher of family violence (wife battering, child abuse, corporal punishment, and the like) found in his research that in physical altercations between husband and wife, the wife was just as likely as the husband to throw the first punch. This is a startling finding when taken by itself. But Straus also learned that regardless of who actually hit first, the wife nearly always wound up being physically injured far more severely than the man. Whoever started the fight, she lost it (Straus & Gelles 1988). In this respect (as well as in certain others), Straus's finding that "women hit first as often as men" is quite misleading when taken by itself. When Straus published his findings, a host of social scientists and feminists protested loudly on the grounds that his research was likely to be misused by those who believe that wife battering is not, in fact, a serious problem. It seemed to suggest that, really, men are no worse in their use of violence than are women. Do researchers have an obligation to try to correct what seem to be misinterpretations of their findings?

Conclusion

Different kinds of research produce different kinds of ethical problems. Most survey research, for instance, creates few if any ethical problems and can even be enjoyable for participants. In fact, researchers from Michigan's Institute for Survey Research interviewed a representative national sample of adults and found that 68% of those who had participated in a survey were somewhat or very interested in participating in another; the more times respondents had been interviewed, the more willing they were to participate again (Reynolds 1979: 56–57). Conversely, some experimental studies in the social sciences that have put people in uncomfortable or embarrassing situations have generated vociferous complaints and years of debate about ethics (Reynolds 1979; Sjoberg 1967).

Encyclopedia Link

Read a brief overview about research ethics.

Research ethics should be based on a realistic assessment of the overall potential for harm and benefit to research subjects. In this chapter, we have presented some basic guidelines, and examples in other chapters suggest applications, but answers aren't always obvious. For example, full disclosure of "what is really going on" in an experimental study is unnecessary if subjects are unlikely to be harmed. In one student observation study on cafeteria workers, for instance, the IRB didn't require consent forms to be signed. The legalistic forms and signatures, they felt, would be more intrusive or upsetting to workers than the very benign and confidential research itself. The committee put the feelings of subjects above the strict requirement for consent.

Ultimately, then, these decisions about ethical procedures are not just up to you, as a researcher, to make. Your university's IRB sets the human subjects protection standards for your institution and will require that researchers—even, in most cases, students—submit their research proposal to the IRB for review. So an institutional committee, following professional codes and guidelines, will guard the ethical propriety of your research; but still, that is an uncertain substitute for your own conscience.

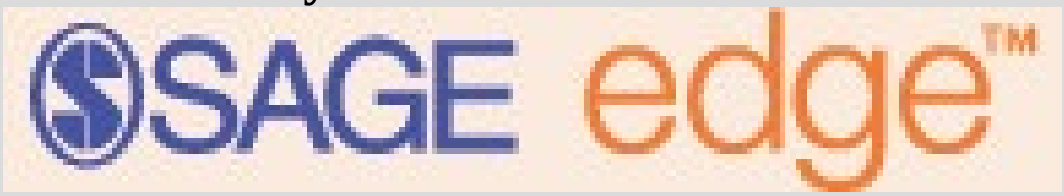
Key Terms

Belmont Report 47
Beneficence 47
Certificate of Confidentiality 56
Confidentiality 56
Debriefing 49
Federal Policy for the Protection of Human Subjects 47
Health Insurance Portability and Accountability Act (HIPAA) 56
Institutional review board (IRB) 47
Justice 47
Nuremberg war crime trials 46
Obedience experiments (Milgram's) 45
Office for Protection From Research Risks, National Institutes of Health 47
Prison simulation study (Zimbardo's) 49
Respect for persons 47
Tearoom Trade 55
Tuskegee syphilis study 46

Highlights

- Milgram’s obedience experiments led to intensive debate about the extent to which deception could be tolerated in psychological research and how harm to subjects should be evaluated.
- Egregious violations of human rights by researchers, including scientists in Nazi Germany and researchers in the Tuskegee syphilis study, led to the adoption of federal ethical standards for research on human subjects.
- The 1979 ***Belmont Report*** developed by a national commission established three basic ethical standards for the protection of human subjects: (1) respect for persons, (2) beneficence, and (3) justice.
- The Department of Health and Human Services adopted the Federal Policy for the Protection of Human Subjects in 1991. The policy requires that every institution seeking federal funding for biomedical or behavioral research on human subjects have an institutional review board to exercise oversight.
- Standards for the protection of human subjects require avoiding harm, obtaining informed consent, avoiding deception except in limited circumstances, and maintaining privacy and confidentiality. Scientific research should maintain high standards for validity and be conducted and reported in an honest and open fashion.
- Effective debriefing of subjects after an experiment can help to reduce the risk of harm caused by the use of deception in the experiment.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Should social scientists be permitted to conduct replications of Milgram's obedience experiments? Zimbardo's prison simulation? Can you justify such research as permissible within the current ASA ethical standards? If not, do you believe that these standards should be altered to permit Milgram-type research?
2. 2. Why does unethical research occur? Is it inherent in science? Does it reflect "human nature"? What makes ethical research more or less likely?
3. Does debriefing solve the problem of subject deception? How much must researchers reveal after the experiment is over, as well as before it begins?

Finding Research

1. The Collaborative Institutional Training Initiative (CITI) offers an extensive online training course in the basics of human subjects protections issues. Go to the public access CITI site (www.citiprogram.org/rcrpage.asp?affiliation=100) and complete the course in social and behavioral research. Write a short summary of what you have learned.
2. 2. The U.S. Department of Health and Human Services maintains extensive resources concerning the protection of human subjects in research. Read several documents that you find on its website (www.hhs.gov/ohrp/), and share your findings in a short report.

Critiquing Research

1. Pair up with one other student and select one of the research articles you have reviewed for other exercises. Criticize the research relative to its adherence to each of the ethical principles for research on human subjects, as well as for the authors' apparent honesty, openness, and consideration of social consequences. Try to be critical but fair. The student with whom you are working should critique the article in the same way but from a generally positive standpoint, defending its adherence to the four guidelines but without ignoring the study's weak points. Together, write a summary of the study's strong and weak points or conduct a debate in class.
2. 2**.** How do you evaluate the current American Sociological Association ethical code? Is it too strict, too lenient, or just about right? Are the enforcement provisions adequate? What provisions could be strengthened?
3. IRB members and the researchers who submit proposals to them must be familiar with a number of key concepts about ethical principles. The interactive exercises "Ethics" lesson at the text's study site will help you learn how to do this.
4. To use these lessons, choose one of the four "Ethics" exercises from the opening menu for the Interactive Exercises. Follow the instructions for entering your answers and responding to the program's comments.
5. Now go to the book's Study Site (edge.sagepub.com/chamblissmssw5e) and choose the "Learning from Journal Articles" option. Read one article based on research involving human subjects. What ethical issues did the research pose, andhow were they resolved? Does it seem that subjects were appropriately protected?

Doing Research

1. List elements in a research plan for the project you envisioned for the “Doing Research” section in Chapter 2 that an IRB might consider to be relevant to the protection of human subjects. Rate each element from 1 to 5, where 1 indicates no more than a minor ethical issue and 5 indicates a major ethical problem that probably cannot be resolved.
2. 2**.** Write one page for the application to the IRB that explains how you will ensure that your research adheres to each relevant standard.

Ethics Questions

1. Read the entire American Sociological Association ***Code of Ethics*** at the ASA website (www.asanet.org/about/ethics.cfm).
2. Discuss the potential challenges in adhering to the ASA's ethical standards in research.

Video Interview Questions

Listen to the researcher interview for Chapter 3 at edge.sagepub.com/chamblissmssw5e.

1. What are the key issues that an institutional review board (IRB) evaluates in a research proposal?
2. What are some challenges that an IRB faces?

Chapter 4 Conceptualization and Measurement

©iStockphoto.com/muratkoc

Learning Objectives

1. Define and distinguish *conceptualization* and *operationalization*.
2. List four different means of operationalizing concepts.
3. Give two examples of constant and two of variable phenomena.
4. Identify the different forms of single questions and response choices.
5. Give examples of the four levels of measurement.
6. Compare the advantages and disadvantages of the three approaches to testing the validity of measures.
7. Define the five methods of evaluating measurement reliability.

Every time you begin to review or design a research study, you will have to answer two questions: (1) What do the main concepts mean in this research? (2) How are the main concepts measured? Both questions must be answered to evaluate the validity of any research. For instance, to study a hypothesized link between religious fundamentalism and terrorism, you may conceptualize terrorism as *nongovernmental political violence* and measure incidents of terrorism by counting, over a 5-year period, the number of violent attacks that have explicit political aims. You will also need to define and measure *religious fundamentalism*—no easy task. What counts? And how should you decide what counts? We cannot make sense of a researcher's study until we know how the concepts were *defined* and *measured.* Nor can we begin our own research until we have defined our concepts clearly and constructed valid measures of them.

In this chapter, we briefly address the issue of conceptualization, or defining your main terms. We then describe measurement sources such as available archive data; questions; observations; and less direct, or unobtrusive, measures. We then discuss the level of measurement reflected in different measures. The final topic is to assess the validity and reliability of these measures. By the chapter's end, you should have a good understanding of measurement, the first of the three legs (measurement, generalizability, and causality) on which a research project's validity rests.

What Do We Have In Mind?

A May 2000 *New York Times* article (Stille 2000) announced that the "social health" of the United States had risen a bit, after a precipitous decline in the 1970s and 1980s. Should we be relieved? Concerned? What, after all, does *social health* mean? The concept of social health means different things to different people. Most agree that it has to do with "things that are not measured in the gross national product" and is supposed to be "a more subtle and more meaningful way of measuring what's important to [people]" (Stille: A19). But until we agree on a definition of social health, we can't decide whether it has to do with child poverty, trust in government, out-of-wedlock births, alcohol-related traffic deaths, or some combination of these or other phenomena.

Conceptualization

A continuing challenge for social scientists, then, rests on the fact that many of our important topics of study (social health, for instance) are not clearly defined things or objects (like trees or rocks) but are abstract concepts or ideas. A **concept** is an image or idea, not a simple object. Some concepts are relatively simple, such as a person's age or sex: Almost everyone would agree what it means to be 14 years old or female. But other concepts are more ambiguous. For instance, if you want to count the number of families in Chicago, what counts as a family? A husband and wife with two biological children living in one house—yes, that's a family. Do cousins living next door count? Cousins living in California? Or maybe the parents are divorced, the children are adopted, or the children are grown. Maybe two women live together with one adopted child and one biological child fathered by a now-absent man. So perhaps "living together" is what defines a family—or is it biology? Or is it a crossing of generations—that is, the presence of adults and children? The particular definition you develop will affect your research findings, and some people probably won't like it whatever you do, but how you define *family* affects your results.

Often social concepts can be used sloppily or even misleadingly. In some years, you may hear that "the economy" is doing well, but even then, many people may be faring badly. Typically in news reports, *the economy* refers to the gross domestic product (GDP)—the total amount of economic activity (value of goods and services, precisely) in the country in a given year. When the GDP goes up, reporters say, "the economy is improving." But that's very different from saying that the average working person makes more money than this person would have 30 years ago—in fact, the average American man makes a little less than 30 years ago, and for women it's close. We could use the concept of *the economy* to refer to the economic well-being of actual people, but that's not typically how it's used.

Defining concepts clearly can be quite difficult because many concepts have several meanings and can be measured in many ways. What is meant, for instance, by the idea of *power*? The classic definition, provided by German sociologist Max Weber (1947/1997: 152), is that power is the ability to meet your goals over the objections of other people. That definition implies that unknown people can be quite powerful, whereas certain presidents of the United

States, very well known, have been relatively powerless. A different definition might equate power to one's official position; in that case, the president of the United States would always be powerful. Or perhaps power is equated with prestige, so famous intellectuals like Albert Einstein would be considered powerful. Or maybe power is defined as having wealth, so that rich people are seen as powerful.

And even if we can settle on a definition, how then do we actually measure power? Should we ask a variety of people if a certain person is powerful? Should we review that person's acts over the last 10 years and see when the person exerted his or her will over others? Should we try to uncover the true extent of the individual's wealth and use that? How about power at a lower level, say, as a member of student government? The most visible and vocal people in your student assembly may be, in fact, quite unpopular and perhaps not very powerful at all—just loud. At the same time, there may be students who are members of no official body whatsoever, but somehow they always get what they want. Isn't that power? From these varied cases, you can see that power can be quite difficult to conceptualize.

Likewise, describing what causes *crime*, or even what causes *theft*, is inherently problematic because the very definition of these terms is spectacularly flexible and indeed forms part of their interest for us. What counts as theft varies dramatically, depending on who is the thief—a next-door neighbor, a sister, or a total stranger wandering through town—and what item is taken: a bottle of water, your watch, a lawn mower, a skirt, your reputation, or $5. Indeed, part of what makes social science interesting is the debates about, for instance, what is a theft or what is crime.

So **conceptualization**—working out what your key terms will mean in your research—is a crucial part of the research process. Definitions need to be explicit. Sometimes conceptualization is easy: "Older men are more likely to suffer myocardial infarction than younger men," or "Career military officers mostly vote for Republican candidates in national elections." Most of the concepts used in those statements are easily understood and easy to measure (gender, age, military status, voting). In other cases, conceptualization is quite difficult: "As people's moral standards deteriorate, the family unit starts to die," or "Intelligence makes you more likely to succeed."

Conceptualization, then, is the process of matching terms (family, sex,

happiness, power) to clarified definitions for them—really, figuring out what are the social “things” you’ll be talking about.

It is especially important to define clearly concepts that are abstract or unfamiliar. When we refer to such concepts as *social capital*, *whiteness*, or*dissonance*, we cannot count on others knowing exactly what we mean. Even experts may disagree about the meaning of frequently used concepts if they base their conceptualizations on different theories. That’s okay. The point is not that there can be only one definition of a concept; rather, we have to specify clearly what we mean when we use a concept, and we should expect others to do the same.

Conceptualization also involves creating concepts, or thinking about how to conceive of the world: What things go together? How do we slice up reality? Smartphones, for instance, may be seen as communication devices, like telephones, radios, telegraphs, or two tin cans connected by a string. Or they can be seen primarily as entertainment devices, like television sets or basketballs. Or they can be conceptualized as being essentially devices for the government to track our activities with—a kind of electronic ankle bracelet that everyone voluntarily carries around. Or they can also be conceived in yet another way: A college administrator we know, seeing students leaving class outside her building, said, “Phones have replaced cigarettes.” She reconceptualized smartphones, seeing them not as communication tools but as something to nervously fiddle with, like cigarettes, chewing gum wrappers, keys on a lanyard, or the split ends of long hair—just “something to do.” In conceptualizing the world, we create the lenses through which we see it.

Research|Social Impact Link

Read about how researchers define and understand the concept of “cool.”

Our point is not that conceptualization problems are insurmountable, but that (1) you need to develop and clearly state what you *mean* by your key concepts, and (2) your measurements will need to be clear and consistent with the definitions you’ve settled on (more on that topic shortly).

Concept: A mental image that summarizes a set of similar observations, feelings, or ideas.

Research That Matters

Social scientists in the United States have conducted many studies of youth gangs during the past one hundred years, but in China, much less research has been conducted about gangs. Only in recent years, as China has rapidly industrialized and urbanized, have social scientists begun to focus attention on Chinese youth gangs.

Vincent Webb, Ling Ren, Jihong "Solomon" Zhao, Ni "Phil" He, and Ineke Haen Marshall decided to study the prevalence of youth gangs in contemporary China, but they realized that the term *youth gang* has not been used in the same way in the two countries. In China, gangs have been defined strictly in terms used in the Chinese legal code, whereas in the United States, the term is used more broadly. Webb et al. asked youth in China and the United States identical questions and found a much lower rate of gang connections in China, but they also discovered that the Chinese word they were using for *gang* did not have the same negative connotation as did the English word. "The Chinese students did not fully understand the concept of gang in the way it is understood in the West due to the cultural and political differences, " they said.

Source: Adapted from Webb, Vincent J., Ling Ren, Jihong "Solomon" Zhao, Ni "Phil" He, and Ineke Haen Marshall. 2011. A comparative study of youth gangs in China and the United States: Definition, offending, and victimization. *International Criminal Justice* Review 21: 225–242.

Conceptualization: The process of specifying what we mean by a term. In deductive research, conceptualization helps translate portions of an abstract theory into testable hypotheses involving specific variables. In inductive research, conceptualization is an important part of the process used to make sense of related observations.

Research in the News

What Is Your Race?

In the 2010 census, 18 million Latinos checked “other” under race. Census researchers and other social science researchers say the measurement error stems from fundamentally different concepts of race. Census data on race and neighborhoods affects the makeup of voting districts and political clout for minority groups. The U.S. Census Bureau continues to alter its line of questioning, and critics argue for new racial thinking.

1. Is the conceptual distinction between race and ethnicity meaningful? Should it be maintained in the census and social science research?
2. Which approach would you recommend? Should social science researchers change their definitions of concepts to reflect popular usage, or should they try to operationalize concepts so that respondents interpret them as desired by the researchers?

News Source: Adapted from Navarro, Mireya. 2012. For many Latinos, racial identity is more culture than color. *New York Times*, January 14: A9.

Variables and Constants

After we define the concepts for a study, we must identify *variables* that correspond to those concepts. For example, we might be interested in what affects students' engagement in their academic work—when they are excited about their studies, when they become eager to learn more, when they enjoy their courses, and so on. We are interested, in other words, in changes in engagement—how and when it varies. Engagement, then, we call a *variable*; it can be high, or it can be low. It's not just a fixed thing. And next, when we try to explain those different levels of student engagement (what causes them), we have to talk about changes in still other things—for instance, in who the teacher is, or what subject teachers offer, or what pedagogical techniques they use. The whole effort to explain something relies on saying, basically, that a change in A causes a change in B. So both A and B have to be changeable things: they must be what scientists call *variables*.

We could use any number of variables to measure engagement: the student's reported interest in classes, teacher evaluations of student engagement, the number of hours spent on homework, or an index summarizing a number of different questions. Any of these variables could show a high or low level of student engagement. If we are to study variation in engagement, we must identify variables to measure that are most pertinent to our theoretical concerns.

Not every concept in a particular study is represented by a variable. In our student engagement study, all of the students *are* students—there is no variation in that. So "student," in this study, is a **constant** (it's always the same), not a variable. You can't explain, for instance, low student engagement in classes by just saying "students are just like that, that's all." If engagement varies, it can only be explained by another variable, not by something that's a constant, or always the case. Or to take a different example, if you studied binge drinking in all-male fraternities, you might believe that the male atmosphere matters. But unless you actually compared them with female groups (sororities, say), gender wouldn't be a variable in your research—because it wouldn't "vary"—and you couldn't actually determine if gender made a difference.

As mentioned, many variables could be used to measure student engagement. Which ones should we select? It's very tempting, and all too common, to simply

try to “measure everything” by including in a study every variable we can think of. We could collect self-reports of engagement, teacher ratings, hours studied per week, pages of essays written for class, number of visits to the library per week, frequency of participation in discussion, times met with professors, and on and on. This haphazard approach will inevitably result in the collection of some useless data and the failure to collect some important data. Instead, we should take four steps:

1. Examine the theories that are relevant to our research question to identify those concepts that would be expected to have some bearing on the phenomenon we are investigating.
2. Review the relevant research literature, and assess the utility of variables used in prior research.
3. Consider the constraints and opportunities for measurement that are associated with the specific setting(s) we will study. Distinguish constants from variables in this setting.
4. Look ahead to our analysis of the data. What role will each variable play in our analysis?

Remember: A few well-chosen variables are better than a barrel full of useless ones.

Constant: A number that has a fixed value in a given situation; a characteristic or value that does not change.

How Will We Know When We've Found It?

Once we have defined our concepts in the abstract—after "conceptualizing"—and we have identified the variables that we want to measure, we must develop our exact measurement procedures; we need to specify the **operations** for measuring the variables we've chosen.

Exhibit 4.1 represents the **operationalization** process for three different concepts. The first researcher defines her concept, binge drinking, and chooses one variable—frequency of heavy episodic drinking—to represent it. This variable is then measured by a specific *indicator*, which in this case will be responses to a single question: "How often within the last 2 weeks did you consume five or more drinks containing alcohol in a row?" (Because "heavy" drinking is defined differently for men and women, the question is phrased as "four or more drinks" for women.). The researcher—moving from left to right on the chart—developed a concept, chose a variable to measure it, then specified the exact *operation* for measuring that variable. Operationalization is the process of turning an abstract concept into a clearly measured variable.

Video Link

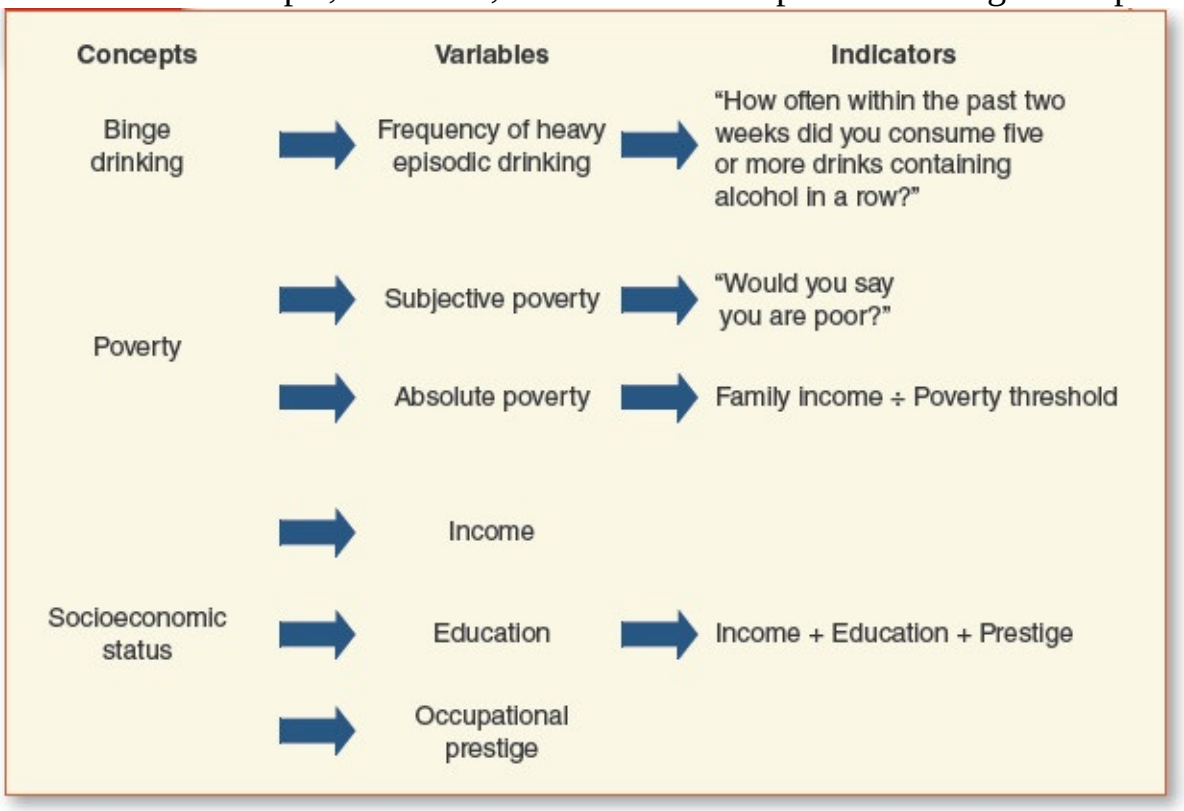
Watch how social scientists find different ways to measure social inequality.

The second researcher defines her concept—poverty— in a more complicated way. She decides that being poor has both subjective and objective components, and both should be measured. (In the research literature, these components are referred to as "subjective" and "absolute" poverty—*absolute* meaning that it's not compared to other people, but to some objective standard.) The variable subjective poverty is then measured (operationalized) with responses to a survey question: "Would you say that you are poor?" Absolute poverty, however, is measured by comparing family income to the poverty threshold. The researcher has operationalized her concept in two different ways.

Finally, the third researcher decides that his concept—socioeconomic status—is multidimensional and should be operationalized by three different variables put together: (1) income, (2) education, and (3) occupational prestige. Only all three

of these combined, he feels, really capture what we mean by social class. So he picks indicators for each, and then puts those all together to provide ratings of a person's social class. Three different operations are used to define social class.

Exhibit 4.1 Concepts, Variables, and Indicators: Operationalizing Concepts



Indicators can be based on activities as diverse as asking people questions, reading judicial opinions, observing social interactions, coding words in books, checking census data tapes, enumerating the contents of trash receptacles, or drawing urine and blood samples. Experimental researchers may operationalize a concept by manipulating its value; for example, to operationalize the concept of exposure to anti-drinking messages, some subjects may listen to a talk about binge drinking, but others do not. In this chapter, we will briefly introduce the operations of using published data, doing content analysis, asking questions, and observing behavior. All of these are covered in more detail later.

The variables and measurement operations chosen for a study should be consistent with the purpose of the research question. Suppose we hypothesize that college students who go abroad for the junior year have a more valuable experience than do those who remain at the college. If our purpose is *evaluation* of different junior-year options, we can operationalize *junior-year programs* by

comparing (1) traditional coursework at home, (2) study in a foreign country, and (3) internships at home that are not traditional college courses. A simple question—for example, asking students in each program, “How valuable do you feel your experience was?”—would help to provide the basis for determining the relative value of these programs. But if our purpose is *explanation*, we would probably want to interview students to learn what features of the different programs made them valuable to find out the underlying dynamics of educational growth.

Time and resource limitations also must be considered when we select variables and devise measurement operations. For many sociohistorical questions (such as “How has the poverty rate varied since 1950?”), census data or other published counts must be used.

A historical question about the types of social bonds among combat troops in wars since 1940 probably requires retrospective interviews with surviving veterans. The validity of the data is lessened by the unavailability of many veterans from World War II and by problems of recall, but direct observation of their behavior during the war is certainly not an option.

Using Available Data

Data can be collected in a wide variety of ways; indeed, much of this book describes different technologies for data collection. But some data are already gathered and ready for analysis (such data will be described in more detail in Chapters 8 and 11). Government reports, for instance, are rich, accessible sources of social science data. Organizations ranging from nonprofit service groups to private businesses also compile a wealth of figures that may be available to some social scientists. Data from many social science surveys are archived and made available for researchers who were not involved in the original survey project.

Before we assume that available data will be useful, we must consider how appropriate they are for our concepts of interest, whether other measures would work better, or whether our concepts can be measured at all with these data. For example, many organizations informally (and sometimes formally) use turnover —that is, how many employees quit each year—as a measure of employee morale (or satisfaction). If turnover is high (or retention rates are low), morale must be bad and needs to be raised. Or so the thinking goes.

But obviously, factors other than morale affect whether people quit their jobs. When a single chicken-processing plant is the only employer in a small town, other jobs are hard to find, and people live on low wages, then turnover may be very low even among miserable workers. In the dot-com companies of the late 1990s, turnover was high—despite amazingly good conditions, salary, and morale—because the industry was so hungry for good workers that companies competed ferociously to attract them. Maybe the concepts *morale* and *satisfaction*, then, can't be measured adequately by the most easily available data (that is, turnover rates).

We also cannot assume that available data are accurate, even when they appear to measure the concept. "Official" counts of homeless persons have been notoriously unreliable because of the difficulty of locating homeless persons on the streets, and government agencies have at times resorted to "guesstimates" by service providers. Even available data for such seemingly straightforward measures as counts of organizations can contain a surprising amount of error. For example, a 1990 national church directory reported 128 churches in a

midwestern county; an intensive search in that county in 1992 located 172 churches (Hadaway, Marler, & Chaves 1993: 744). Still, when legal standards, enforcement practices, and measurement procedures have been considered, comparisons among communities become more credible.

However, such adjustments may be less necessary when the operationalization of a concept is seemingly unambiguous, as with the homicide rate: After all, dead is dead, right? And when a central authority imposes a common data collection standard, as with the FBI's *Uniform Crime Reports*, data become more comparable across communities. But even here, careful review of measurement operations is still important because (for instance) procedures for classifying a death as a homicide can vary between jurisdictions and over time.

Researcher Interview Link

Watch a Researcher Explain the GSS and how we use it.

Another rich source of already-collected data is survey data sets archived and made available to university researchers by the Inter-University Consortium for Political and Social Research (1996). One of its most popular survey datasets is the General Social Survey (GSS). The GSS is administered regularly by the National Opinion Research Center (NORC) at the University of Chicago to a sample of more than 1,500 Americans (annually until 1994; biennially since then). GSS questions vary from year to year, but an unchanging core of questions includes measures of political attitudes, occupation and income, social activities, substance abuse, and many other variables of interest to social scientists. College students can easily use this data set to explore a wide range of interesting topics. However, when surveys are used in this way, after the fact, researchers must carefully evaluate the survey questions. Are the available measures sufficiently close to the measures needed that they can be used to answer the new research question?

Operation: A procedure for identifying or indicating the value of cases on a variable.

Operationalization: The process of specifying the operations that will indicate the value of cases on a variable.

Content Analysis

One particular method for using available data is **content analysis**, a method for systematically analyzing and making inferences from text (Weber 1985: 9). You can think of a content analysis as a survey of "documents," ranging from newspapers, books, or TV shows to persons referred to in other communications, themes expressed in government documents, or propositions made in tape-recorded debates. Words or other features of these units are then coded to measure the variables involved in the research question (Weber). As a simple example of content analysis, you might look at a variety of women's magazines from the past 25 years and count the number of articles in each year devoted to various topics, such as makeup, weight loss, relationships, sex, and so on. You might count the number of articles on different subjects as a measure of the media's emphasis on women's anxiety about these issues and see how that emphasis (i.e., the number of articles) has increased or decreased during the past quarter century. At the simplest level, you could code articles by whether key words *(fat, weight, pounds*, etc.) appeared in the titles.

After coding procedures are developed, their reliability should be assessed by comparing different coders' results for the same variables. Computer programs for content analysis can be used to enhance reliability (Weitzman & Miles 1994). The computer is programmed with certain rules for coding text so that these rules will be applied consistently. We describe content analysis in detail in Chapter 11.

Journal Link

Read about a study using a content analysis design that looks at gender, race, and sports.

Content analysis: A research method for systematically analyzing and making inferences from text.

Constructing Questions

Asking people questions is the most common, and probably most versatile, operation for measuring social variables. Do you play on a varsity team? What is your major? How often, in a week, do you go out with friends? How much time do you spend on schoolwork? Most concepts about individuals can be measured with such simple questions. In this section, we introduce some options for writing questions, explain why single questions can sometimes be inadequate measures, and then examine the use of multiple questions to measure a concept.

In principle, questions, asked perhaps as part of a survey, can be a straightforward and efficient means by which to measure individual characteristics, facts about events, level of knowledge, and opinions of any sort. In practice, though, survey questions can easily result in misleading or inappropriate answers. All questions proposed for a survey must be screened carefully for their adherence to basic guidelines and then tested and revised until the researcher feels some confidence that they will be clear to the intended respondents (Fowler 1995). Some variables may prove to be inappropriate for measurement with any type of question. We have to recognize that memories and perceptions of the events about which we might like to ask can be limited.

Specific guidelines for reviewing questions are presented in Chapter 7; here, our focus is on the different types of survey questions.

Single Questions

Measuring variables with single questions is very popular. Public opinion polls based on answers to single questions are reported frequently in newspaper articles and TV newscasts: Do you favor or oppose U.S. policy in Iraq? If you had to vote today, for which candidate would you vote? Social science surveys also rely on single questions to measure many variables: Overall, how satisfied are you with your job? How would you rate your current health?

Researcher Interview Link

Watch a researcher describe how to construct questions.

Single questions can be designed with or without explicit response choices. The question that follows is a **closed-ended**, or **fixed-choice**, **question** because respondents are offered explicit responses from which to choose. It has been selected from the Core Alcohol and Drug Survey distributed by the Core Institute, Southern Illinois University, for the Fund for the Improvement of Postsecondary Education (FIPSE) Core Analysis Grantee Group (Presley, Meilman, & Lyerla 1994).

Compared with other campuses with which you are familiar, this campus's use of alcohol is . . . (Mark one)

_____*Greater than other campuses*

_____*Less than other campuses*

_____*About the same as other campuses*

Most surveys of a large number of people contain primarily fixed-choice questions, which are easy to process with computers and analyze with statistics. However, fixed-response choices can obscure what people really think, unless the choices are designed carefully to match the range of possible responses to the question.

Most important, response choices should be **mutually exclusive** and **exhaustive**, so that respondents can each find *one and only one* choice that applies to them (unless the question is of the "Check all that apply" variety). To make response choices exhaustive, researchers may need to offer at least one option with room for ambiguity. For example, a questionnaire asking college students to indicate their school status should not use freshman, sophomore, junior, senior, and graduate student as the only response choices. Most campuses also have students in a "special" category, so you might add "Other (please specify)" to the five fixed responses to this question. If respondents do not find a response option that corresponds to their answer to the question, they may skip the question entirely or choose a response option that does not indicate what they are really thinking.

Researchers who study small numbers of people often use **open-ended questions**, which don't have explicit response choices and allow respondents to write in their answers. The next question is an open-ended version of the earlier

fixed-choice question:

> *How would you say alcohol use on this campus compares to that on other campuses?*

An open-ended format is preferable when the full range of responses cannot be anticipated, especially when questions have not been used previously in surveys or when questions are asked of new groups. Open-ended questions also can allow clear answers when questions involve complex concepts. In the previous question, for instance, “alcohol use” may cover how many students drink, how heavily they drink, if the drinking is public or not, if it affects levels of violence on campus, and so on.

Just like fixed-choice questions, open-ended questions should be reviewed carefully for clarity before they are used. For example, if respondents are asked, “When did you move to Boston?” they might respond with a wide range of answers: “In 1987.” “After I had my first child.” “When I was 10.” “20 years ago.” Such answers would be very hard to compile. To avoid such ambiguity, rephrase the question to clarify the form of the answer; for instance, “In what year did you move to Boston?” Or provide explicit response choices (Center for Survey Research 1987).

Encyclopedia Link

Read about why and how open-ended questions should be implemented.

Closed-ended (fixed-choice) question: A survey question that provides preformatted response choices for the respondent to circle or check.

Mutually exclusive: A variable’s attributes (or values) are mutually exclusive when every case can be classified as having only one attribute (or value).

Exhaustive: Every case can be classified as having at least one attribute (or value) for the variable.

Open-ended question: A survey question to which respondents reply in their own words, either by writing or by talking.

Indexes and Scales

When several questions are used to measure one concept, the responses may be combined by taking the sum or average of responses. A composite measure based on this type of sum or average is termed an **index**. The idea is that idiosyncratic variation in response to particular questions will average out, so that the main influence on the combined measure will be the concept that all the questions focus on. In addition, the index can be considered a more complete measure of the concept than can any one of the component questions.

Creating an index is not just a matter of writing a few questions that seem to focus on a concept. Questions that seem to you to measure a common concept might seem to respondents to concern several different issues. The only way to know that a given set of questions actually does form an index is to administer the questions to people like those you plan to study. If a common concept is being measured, people's responses to the different questions should display some consistency.

Because of the popularity of survey research, indexes already have been developed to measure many concepts, and some of these indexes have proven to be reliable in a range of studies. Usually it is much better to use such an index than it is to try to form a new one. Use of a preexisting index both simplifies the work of designing a study and facilitates the comparison of findings from other studies.

The questions in Exhibit 4.2 represent a short form of an index used to measure depression; it is called the Center for Epidemiologic Studies Depression Index (CES-D). Many researchers in different studies have found that these questions form a reliable index. Note that each question concerns a symptom of depression. People may well have one particular symptom without being depressed; for example, persons who have been suffering from a physical ailment may say that they have a poor appetite. By combining the answers to questions about several symptoms, the index reduces the impact of this idiosyncratic variation. (This set of questions uses what is termed a *matrix* format, in which a series of questions that concern a common theme are presented together with the same response choices.)

Exhibit 4.2 Examples of Indexes: Short Form of the Center for Epidemiologic Studies (CES-D) and "Negative Outlook" Index

At any time during the past week . . . (Circle one response on each line)	Never	Some of the Time	Most of the Time
a. Was your appetite so poor that you did not feel like eating?	1	2	3
b. Did you feel so tired and worn out that you could not enjoy anything?	1	2	3
c. Did you feel depressed?	1	2	3
d. Did you feel unhappy about the way your life is going?	1	2	3
e. Did you feel discouraged and worried about your future?	1	2	3
f. Did you feel lonely?	1	2	3
Negative outlook			
How often was each of these things true during the past week? (Circle one response on each line)	**A Lot, Most, or All of the Time**	**Sometimes**	**Never or Rarely**
a. You felt that you were just as good as other people.	0	1	2
b. You felt hopeful about the future.	0	1	2
c. You were happy.	0	1	2
d. You enjoyed life.	0	1	2

Source: Hawkins, Daniel N., Paul R. Amato, and Valarie King. 2007. Nonresident father involvement and adolescent well-being: Father effects or child effects? American Sociological Review 72: 990.

Usually an index is calculated by simply averaging responses to the questions, so that every question counts equally. But sometimes, either intentionally by the researcher or by happenstance, questions on an index arrange themselves in a kind of hierarchy in which an answer to one question effectively provides answers to others. For instance, a person who supports abortion on demand almost certainly supports it in cases of rape and incest as well. Such questions form a **scale**. In a scale, we give different weights to the responses to different questions before summing or averaging the responses. Responses to one question might be counted two or three times as much as responses to another. For example, based on Christopher Mooney and Mei Hsien Lee's (1995)

research on abortion law reform, a scale to indicate support for abortion might give a 1 to agreement that abortion should be allowed “when the pregnancy results from rape or incest” and a 4 to agreement with the statement that abortion should be allowed “whenever a woman decides she wants one.” A 4 rating is much stronger, in that anyone who gets a 4 would probably agree to all lower-number questions as well.

Index: A composite measure based on summing, averaging, or otherwise combining the responses to multiple questions that are intended to measure the same concept.

Scale: A composite measure based on combining the responses to multiple questions pertaining to a common concept after these questions are differentially weighted, such that questions judged on some basis to be more important for the underlying concept contribute more to the composite score.

Making Observations

Asking questions, then, is one way to operationalize, or measure, a variable. *Observations* can also be used to measure characteristics of individuals, events, and places. The observations may be the primary form of measurement in a study, or they may supplement measures obtained through questioning.

Direct observations can be used as indicators of some concepts. For example, Albert J. Reiss Jr. (1971) studied police interaction with the public by riding in police squad cars, observing police–citizen interactions, and recording the characteristics of the interactions on a form. Notations on the form indicated such variables as how many police–citizen contacts occurred, who initiated the contacts, how compliant citizens were with police directives, and whether police expressed hostility toward the citizens.

Often, observations can supplement what is initially learned from interviews or survey questions, putting flesh on the bones of what is otherwise just a verbal self-report. In Daniel Chambliss's (1996) book, *Beyond Caring*, a theory of the nature of moral problems in hospital nursing that was originally developed through interviews was expanded with lessons learned from observations. Chambliss found, for instance, that in interviews, nurses described their daily work as exciting, challenging, dramatic, and often even heroic. But when Chambliss himself sat for many hours and watched nurses work, he found that their daily lives were rather humdrum and ordinary, even to them. Occasionally, there were bursts of energetic activity and even heroism—but the reality of day-to-day nursing was far less exciting than interviews would lead one to believe. Indeed, Chambliss modified his original theory to include a much broader role for routine in hospital life.

Direct observation is often the method of choice for measuring behavior in natural settings, as long as it is possible to make the requisite observations. Direct observation avoids the problems of poor recall and self-serving distortions that can occur with answers to survey questions. It also allows measurement in a context that is more natural than an interview. But observations can be distorted, too. Observers do not see or hear everything, and their own senses and perspectives filter what they do see. Moreover, in some situations, the presence of an observer may cause people to act differently from the way they would

otherwise (Emerson 1983). If you set up a video camera in an obvious spot on campus to monitor traffic flows, you may well change the flow—just because people will see the camera and avoid it (or come over to make faces). We will discuss these issues in more depth in Chapter 9, but it is important to begin to consider them whenever you read about observational measures.

Combining Measurement Operations

The choice of a particular measurement method—questions, observations, archives, and the like—is often determined by available resources and opportunities, but measurement is improved if this choice also considers the particular concept or concepts to be measured. Responses to questions such as "How socially adept were you at the party?" or "How many days did you use sick leave last year?" are unlikely to provide valid information on shyness or illness. Direct observation or company records may work better. Conversely, observations at cocktail parties may not fully answer our questions about why some people are shy; we may just have to ask people. Or if a company keeps no record of sick leave, we may have to ask direct questions and hope for accurate memories. Every choice of a measurement method entails some compromise between the perfect and the possible.

Triangulation—the use of two or more different measures of the same variable—can strengthen measurement considerably (Brewer & Hunter 1989: 17). When we achieve similar results with different measures of the same variable, particularly when they are based on such different methods as survey questions and field-based observations, we can be more confident of the validity of each measure. In surveys, for instance, people may say that they would return a lost wallet they found on the street. But field observation may prove that in practice, many succumb to the temptation to keep the wallet. The two methods produce different results. In a contrasting example, post-combat interviews of U.S. soldiers in World War II found that most GIs never fired their weapons in battle, and the written, archival records of ammunition resupply patterns confirmed this interview finding (Marshall 1947/1978). If results diverge when using different measures, it may indicate that we are sustaining more measurement error than we can tolerate.

Divergence between measures could also indicate that each measure actually operationalizes a different concept. An interesting example of this interpretation of divergent results comes from research on crime. Crime statistics are often inaccurate measures of actual crime; what gets reported to the police and shows up in official statistics is not at all the same thing as what happens according to victimization surveys (in which random people are asked if they have been a crime victim). Social scientists generally regard victim surveys as a more valid

measure of crime than police-reported crime. We know, for instance, that rape is a dramatically *underreported* crime, with something like 4 to 10 times the number of rapes occurring as are reported to police. But auto theft is an *overreported* crime: More auto thefts are reported to police than actually occur. This may strike you as odd, but remember that almost everyone who owns a car also owns car insurance; if the car is stolen, the victim will definitely report it to the police to claim the insurance. Plus, some other people might report cars stolen when they haven't been because of the financial incentive. (By the way, insurance companies are quite good at discovering this scam, so it's a bad way to make money.)

Researcher Interview Link

Watch a researcher discuss how to receive adequate measurements.

Murder, however, is generally reported to police at roughly the same rate at which it actually occurs (i.e., official police reports generally match victim surveys). When someone is killed, it's very difficult to hide the fact: A body is missing, a human being doesn't show up for work, people find out. At the same time, it's very hard to pretend that someone was murdered when the person wasn't murdered. There he or she is, still alive, in the flesh. Unlike rape or auto theft, there are no obvious incentives for either underreporting or overreporting murders. The official rate is generally valid.

So if you can, it's best to use multiple measures of the same variable; that way, each measure helps to check the validity of the others.

Triangulation: The use of multiple methods to study one research question.

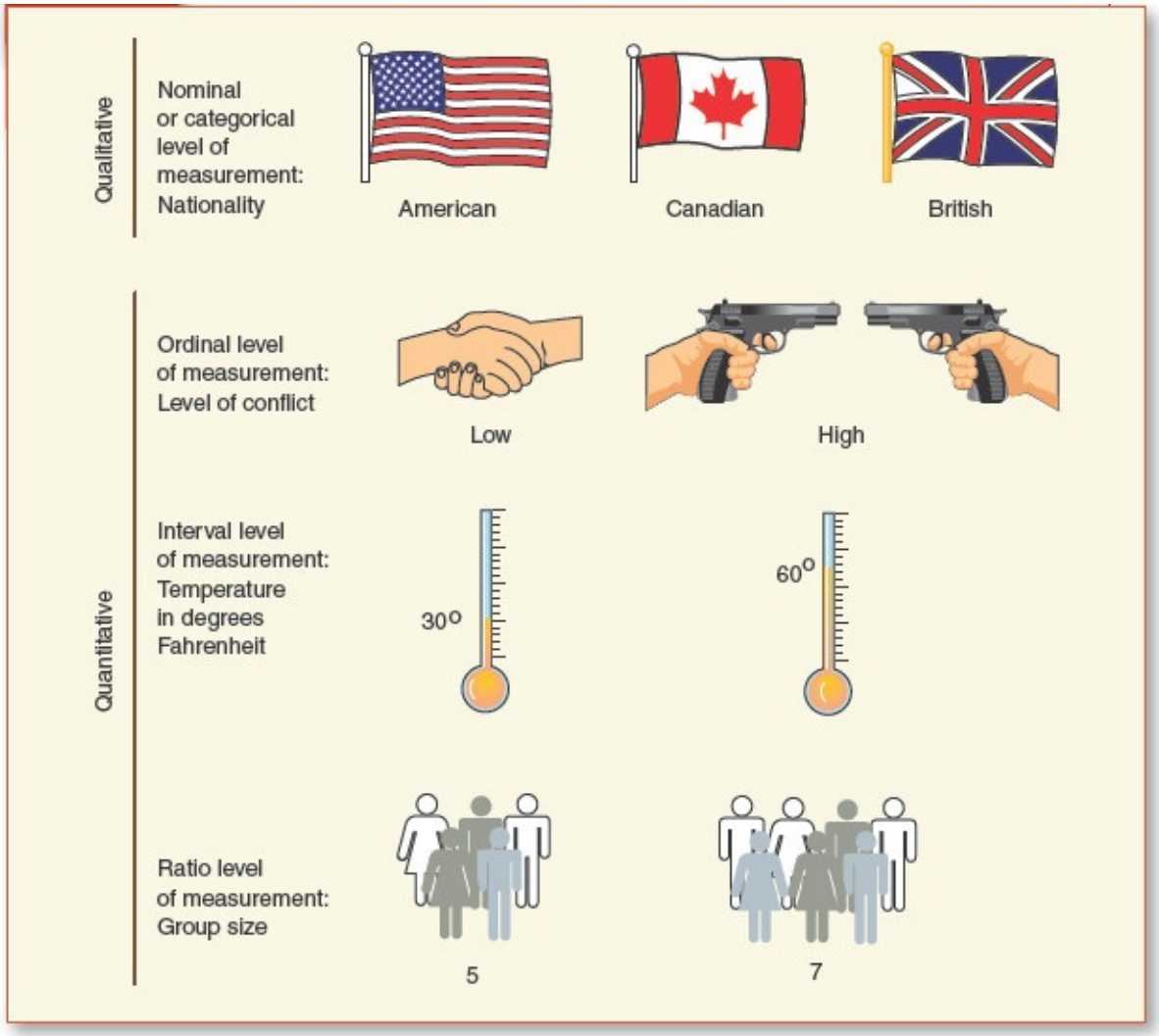
How Much Information Do We Really Have?

There are many ways of collecting information, or different *operations* for gathering data: asking questions, using previously gathered data, analyzing texts, and so on. Some of this data contains mathematically detailed information; it represents a higher level of measurement. There are four **levels of measurement**: (1) nominal, (2) ordinal, (3) interval, and (4) ratio. Exhibit 4.3 depicts the differences among these four levels.

Nominal Level of Measurement

The **nominal level of measurement** identifies variables whose values have no mathematical interpretation; they vary in kind or quality but not in amount. *State* (referring to the United States) is one example. The variable has 50 attributes (or categories or qualities), but none of them is more *state* than another. They're just different. *Religious affiliation* is another nominal variable, measured in categories: Christian, Muslim, Hindu, Jewish, and so on. *Nationality*, *occupation*, and*region of the country* are also measured at the nominal level. A person may be Spanish or Portuguese, but one nationality does not represent more nationality than another—just a different nationality (Exhibit 4.3). A person may be a doctor or a truck driver, but one does not represent three units "more occupation" than the other. Of course, more people may identify themselves as being of one nationality than of another, or one occupation may have a higher average income than another occupation, but these are comparisons involving variables other than *nationality* or *occupation* themselves.

Exhibit 4.3 Levels of Measurement



Journal Link

Read an article that uses nominal variables to help assess community networks.

Although the attributes of nominal variables do not have a mathematical meaning, they must be assigned to cases with great care. The attributes we use to measure, or categorize, cases must be mutually exclusive and exhaustive:

Interactive Exercises

Levels of Measurement

- A variable’s attributes or values are mutually exclusive if every case can have only one attribute.
- A variable’s attributes or values are exhaustive when every case can be classified into one of the categories.

When a variable’s attributes are mutually exclusive and exhaustive, every case corresponds to one—and only one—attribute.

Level of measurement: The mathematical precision with which the values of a variable can be expressed. The nominal level of measurement, which is qualitative, has no mathematical interpretation; the quantitative levels of measurement—ordinal, interval, and ratio—are progressively more precise mathematically.

Nominal level of measurement: Variables whose values have no mathematical interpretation; they vary in kind or quality but not in amount.

Ordinal Level of Measurement

The first of the three quantitative levels is the **ordinal level of measurement**. At this level, you specify only the order of the cases in *greater than* and *less than* distinctions. At the coffee shop, for example, you might choose between a small, medium, or large cup of decaf—that's ordinal measurement.

The properties of variables measured at the ordinal level are illustrated in Exhibit 4.3 by the contrast between the levels of conflict in two groups. The first group, symbolized by two people shaking hands, has a low level of conflict. The second group, symbolized by two people pointing guns at each other, has a high level of conflict. To measure conflict, we could put the groups "in order" by assigning 1 to the low-conflict group and 2 to the high-conflict group, but the numbers would indicate only the relative position, or order, of the cases.

As with nominal variables, the different values of a variable measured at the ordinal level must be mutually exclusive and exhaustive. They must cover the range of observed values and allow each case to be assigned no more than one value.

Video Link

Watch more about levels of measurement.

Ordinal level of measurement: A measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.

Interval Level of Measurement

At the **interval level of measurement**, numbers represent fixed measurement units but have no absolute zero point. This level of measurement is represented in Exhibit 4.3 by the difference between two Fahrenheit temperatures. Note, for example, that 60 degrees is 30 degrees higher than 30 degrees, but 60 is not "twice as hot" as 30. Why not? Because heat does not "begin" at 0 degrees on the Fahrenheit scale. The numbers can therefore be added and subtracted, but ratios of them (2 to 1 or "twice as much") are not meaningful. There are thus few true interval-level measures in the social sciences; most are ratio-level, because they have zero points.

Sometimes, though, social scientists will create indexes by combining responses to a series of variables measured at the ordinal level and then treat these indexes as interval-level measures. An index of this sort could be created with responses to the Core Institute's questions about friends' disapproval of substance use (Exhibit 4.4). The survey has 13 questions on the topic, each of which has the same three response choices. If "Don't disapprove" is valued at 1, "Disapprove" is valued at 2, and "Strongly disapprove" is valued at 3, the summed index of disapproval would range from 13 to 39. A score of 20 could be treated as if it were 4 more units than a score of 16. Or the responses could be averaged to retain the original 1 to 3 range.

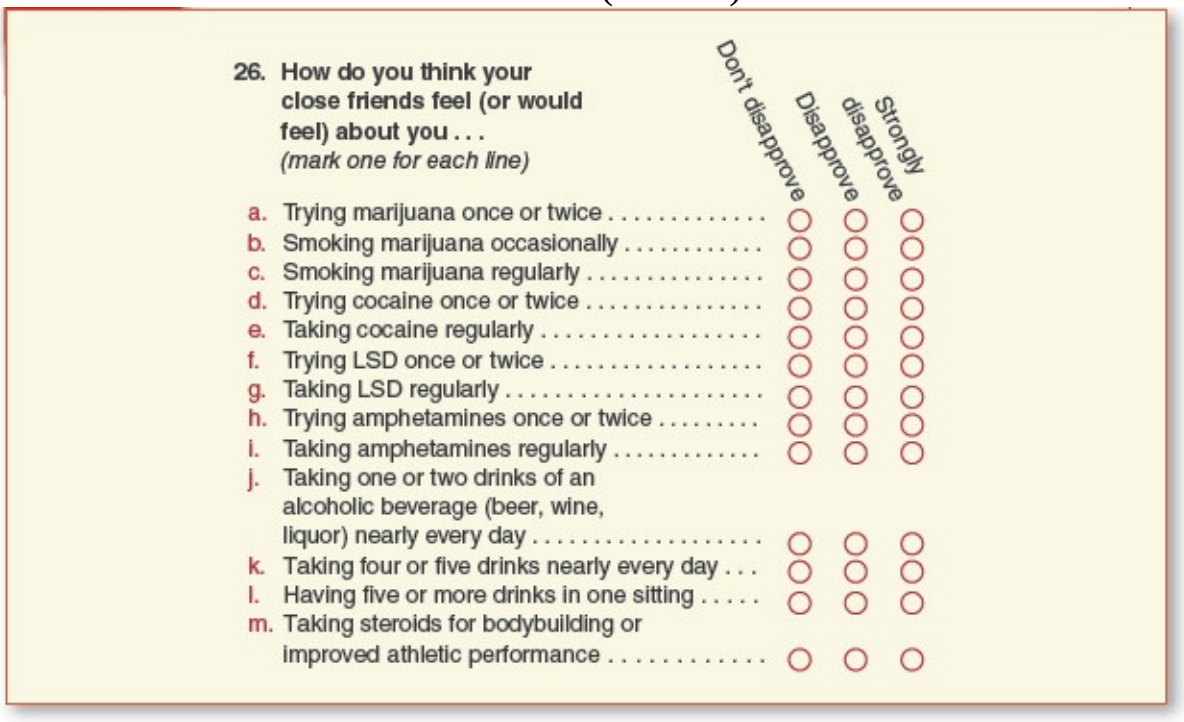
Interval level of measurement: A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.

Ratio Level of Measurement

A **ratio level of measurement** represents fixed measuring units with an absolute zero point. Zero, in this situation, means absolutely no amount of whatever the variable indicates. On a ratio scale, 10 is 2 points higher than 8 and is also 2 times as great as 5. Ratio numbers can be added and subtracted, and because the numbers begin at an absolute zero point, they can also be multiplied and divided (so ratios can be formed between the numbers).

For example, people's ages can be represented by values ranging from 0 years (or some fraction of a year) to 120 or more. A person who is 30 years old is 15 years older than someone who is 15 years old (30 – 15 = 15) and is also twice as old as that person (30/15 = 2). Of course, the numbers also are mutually exclusive and exhaustive, so that every case can be assigned one and only one value. Age (in years) is clearly a ratio-level measure.

Exhibit 4.4 Ordinal Measures: Core Alcohol and Drug Survey. Responses could be combined to create an interval scale (see text).



26. How do you think your close friends feel (or would feel) about you . . .
(mark one for each line)

	Don't disapprove	Disapprove	Strongly disapprove
a. Trying marijuana once or twice	○	○	○
b. Smoking marijuana occasionally	○	○	○
c. Smoking marijuana regularly	○	○	○
d. Trying cocaine once or twice	○	○	○
e. Taking cocaine regularly	○	○	○
f. Trying LSD once or twice	○	○	○
g. Taking LSD regularly	○	○	○
h. Trying amphetamines once or twice	○	○	○
i. Taking amphetamines regularly	○	○	○
j. Taking one or two drinks of an alcoholic beverage (beer, wine, liquor) nearly every day	○	○	○
k. Taking four or five drinks nearly every day	○	○	○
l. Having five or more drinks in one sitting	○	○	○
m. Taking steroids for bodybuilding or improved athletic performance	○	○	○

Source: Core Institute 1994. Core alcohol and drug survey. Carbondale, IL:

Core Institute.

Exhibit 4.3 displays an example of a variable measured at the ratio level. The number of people in the first group is 5, and the number in the second group is 7. The ratio of the two groups' sizes is then 1.4, a number that mirrors the relationship between the sizes of the groups. Note that there does not actually have to be any "group" with a size of zero; what is important is that the numbering scheme begins at an absolute zero—in this case, the absence of any people.

Ratio level of measurement: A measurement of a variable in which the numbers indicating the variable's values represent fixed measuring units *and* an absolute zero point.

Comparison of Levels of Measurement

Exhibit 4.5 summarizes the types of comparisons that can be made with different levels of measurement, as well as the mathematical operations that are legitimate with each. All four levels of measurement allow researchers to assign different values to different cases. All three quantitative measures allow researchers to rank cases in order.

Researchers choose levels of measurement in the process of operationalizing variables; the level of measurement is not inherent in the variable itself. Many variables can be measured at different levels with different procedures. Age can be measured as *young* or *old*; as 0 to 10, 11 to 20, 21 to 30, and so on; or as 1, 2, or 3 years old. We could gather the data by asking people their age, by having an observer guess ("Now *there's* an old guy!"), or by searching through hospital records for exact dates and times of birth. Any of these approaches could work, depending on our research goals.

Usually, though, it is a good idea to measure variables at the highest level of measurement possible. The more information available, the more ways we have to compare cases. We also have more possibilities for statistical analysis with quantitative than with qualitative variables. Even if your primary concern is only to compare teenagers to young adults, you should measure age in years rather than in categories; you can always combine the ages later into categories corresponding to *teenager* and *young adult.*

Exhibit 4.5 Properties of Measurement Levels

Examples of Comparison Statements	Appropriate Math Operations	Relevant Level of Measurement			
		Nominal	Ordinal	Interval	Ratio
A is equal to (not equal to) B	= (≠)	✓	✓	✓	✓
A is greater than (less than) B	> (<)		✓	✓	✓
A is three more than (less than) B	+ (–)			✓	✓
A is twice (half) as large as B	× (/)				✓

Be aware, however, that other considerations may preclude measurement at a high level. For example, many people are very reluctant to report their exact incomes, even in anonymous questionnaires. So asking respondents to report

their income in categories (such as less than $10,000, $10,000–$19,999, $20,000–$29,999, and so on) will elicit more responses, and thus more valid data, than will asking respondents for their income in dollars.

Careers and Research

Dana Hunt, PhD, Principal Scientist

Source: Dana Hunt

In the study site video for this chapter, Dana Hunt discusses two of the many lessons she has learned about measurement in a decades-long career in social research. Hunt received her BA in sociology from Hood College in Pennsylvania and then earned her PhD in sociology at the University of Pennsylvania. After teaching at Hood for several years, she took an applied research position at National Development and Research Institutes (NDRI) in New York City. NDRI's description on its website gives you an idea of what drew the attention of a talented young social scientist.

Founded in 1967, NDRI is a nonprofit research and educational organization dedicated to advancing scientific knowledge in the areas of drug and alcohol abuse, treatment, and recovery; HIV, AIDS and HCV [hepatitis C virus]; therapeutic communities; youth at risk; and related areas of public health, mental health, criminal justice, urban problems, prevention, and epidemiology.

Hunt moved from New York to the Boston area in 1990, where she is now a principal scientist at Abt Associates, Inc., in Cambridge. Abt's website description conveys the scope of the research projects the company directs.

Abt Associates applies scientific research, consulting, and technical assistance expertise on a wide range of issues in social, economic, and health policy; international development; clinical trials; and registries. One of the largest for-profit government and business research and consulting firms in the world, Abt Associates delivers practical, measurable, high-value-added results.

Two of Hunt's major research projects in recent years are the nationwide Arrestee Drug Abuse Monitoring Program for the Office of National Drug Control Policy and a study of prostitution and sex trafficking demand reduction for the National Institute of Justice.

Did We Measure What We Wanted to Measure?

A good measurement needs to be both *valid* and *reliable*. Measurement validity, as we've discussed in Chapter 1, means that an operation should accurately measure what it's supposed to. There are ways to check on validity, as explained later. A measurement should also be reliable, in the sense of producing consistent answers.

Measurement Validity

A good measure of a person's age is the current year minus the year given on that person's birth certificate. Very probably, the resulting number accurately represents the person's age. This would be a *valid* measure. A less valid measure would be for the researcher to ask the person (who may lie or forget) or for the researcher to simply guess.

Audio Link

Listen to an issue regarding measurement validity.

Measurement validity can be assessed in several ways: (1) face validation, (2) criterion validation, and (3) construct validation.

Face Validity

Researchers apply the term **face validity** to the confidence gained from careful inspection of a concept to see if it is appropriate "on its face." More precisely, we can say that a measure has face validity if it obviously pertains to the meaning of the concept being measured more than to other concepts (Brewer & Hunter 1989: 131). For example, a count of the number of drinks people have consumed in the past week would be a measure of their alcohol consumption that has face validity. It just seems obviously appropriate.

Although every measure should initially be inspected in this way, face validity is not scientifically convincing. Face validity helps, but often not much. For instance, let's say that Sara is having some worries about her boyfriend, Jeremy. She wants to know if he loves her. So she asks him (her measurement!), "Jeremy, do you really love me?" He replies, "Sure, baby, you know I do." That's face validity; she doesn't think he's a shameless liar. And yet Jeremy routinely goes out with other women, only calls Sara once every 3 weeks, and isn't particularly nice to her when they do go out. His answer that he loves her has face validity, but Sara should probably look for other validating measures—for instance, how he actually treats her and their relationship.

Face validity: The type of validity that exists when an inspection of items used to measure a concept suggests that they are appropriate “on their face.”

Criterion Validity

Much stronger than face validity is **criterion validity.** Criterion validity is established when the results from one measure match those obtained with a more direct or an already-validated measure of the same phenomenon (the *criterion*). A measure of blood-alcohol concentration, for instance, could be the criterion for validating a self-report measure of drinking. In other words, if Jason says he hasn’t been drinking, we establish criterion validity by giving him a Breathalyzer test. Observations of drinking by friends or relatives could also, in some limited circumstances, serve as a criterion for validating a self-report.

Criterion validity is established when a more direct measure of the phenomenon regularly points to the same answer as the measure we seek to validate. A store might validate a written test of sales ability comparing test scores to peoples’ actual sales performance. Or, a measure of walking speed based on mental counting might be validated with a stopwatch. Predicting scores on a criterion measured in the future can sometimes validate a measure—for instance, if SAT scores accurate predict college grades, that would validate the SAT.

For many concepts social scientists are interested in—for instance, human emotions—it’s difficult to find a well-established criterion. Yes, if you and your roommate are together every evening, you can actually count the beers he seems to be drinking every night. You definitely know about his drinking. But if we want to measure his feelings of social awkwardness or exclusion, what direct indicator could serve as a criterion? How do you really know if he’s feeling bad? A tax return can validate self-reported income, but what would you use to measure misery?

Criterion validity: The type of validity that is established by comparing the scores obtained on the measure being validated to those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

Construct Validity

Measurement validity also can be established by relating a measure to other

measures, as specified in a theory. Different parts of a theory should "hang together"; if they do, this helps to validate the measures.This approach, known as **construct validity**, is commonly used in social research when no clear criterion exists for validation purposes.

An historically famous example of construct validity is provided by the work of Theodor W. Adorno, Nevitt Sanford, Else Frenkel-Brunswik, and Daniel Levinson (1950) in their book *The Authoritarian Personality.* Adorno and his colleagues, working in the United States and Germany immediately after World War II, were interested in a question that troubled much of the world during the 1930s and 1940s: Why were so many people attracted to Nazism and to its Italian and Japanese fascist allies? Hitler was not an unpopular leader in Germany. In fact, in January 1933, he came to power by being elected chancellor (something like president) of Germany, although some details of the election were a bit suspicious. Millions of people supported him enthusiastically. Why did so many Germans during the 1930s come to nearly worship Adolf Hitler and believe strongly in his program—which proved, of course, to be so disastrous for Europe and the rest of the world? The Adorno research group proposed the existence of what they called an "authoritarian personality," a type of person who would be drawn to a dictatorial leader of the Hitler type. Their key concept, then, was *authoritarianism.*

But of course, there's no such "thing" as authoritarianism; it's not like a tree, something you can look at. It's a *construct,* an idea that we use to help make sense of the world. To establish construct validity of this idea, the researchers created a number of different scales made up of interview questions. One scale was called the "anti-Semitism" scale, in which hatred of Jews was measured. Another measure was a "fascism" scale, measuring a tendency toward favoring a militaristic, nationalist government. Another was the "political and economic conservatism" scale, and so on. Adorno and his colleagues interviewed lots of Germans and found that high scores on these different scales tended to correlate; a person who scored high on one tended to score high on the others. Hence, they determined that the authoritarian personality was a legitimate construct. The idea of authoritarianism, then, was validated through construct validity. Simultaneous high scores on them validated the idea of authoritarianism.

Perhaps a simpler example is establishing the validity of the Addiction Severity Index (ASI). A. Thomas McLellan and his associates (1985) compared subject scores on the ASI with a number of indicators that they felt, from prior research,

should be related to substance abuse: medical problems, employment problems, legal problems, family problems, and psychiatric problems. The researchers could not use a criterion validation approach because they did not have a more direct measure of abuse, such as laboratory test scores or observer reports. However, their extensive research on the subject had given them confidence that these sorts of problems were all related to substance abuse, and, indeed, they found that individuals with higher ASI ratings tended to have more problems in each of these areas.

Both criterion and construct validation, then, compare scores on one measure to scores on other measures that are predicted to be related. Distinguishing the two forms (criterion and construct) matters less than thinking clearly about the comparison measures and whether they actually represent different views of the same phenomenon. For example, correspondence between scores on two different self-report measures of alcohol use (“Are you a heavy drinker?” “How many drinks would say you have in a week?”) is a weak indicator of measurement validity. A person just reports in two different ways how much she drinks; of course the two will be related. A self-report measure validated with an observer-based measure of substance use would be much stronger. The subject (1) reports how much she drinks, and then (2) an observer reports on the subject’s drinking. If the results match up, it’s strong evidence of validity. Thus, criterion validation can be considered as a more rigorous form of validation, if there is a clear criterion to use.

Construct validity: The type of validity that is established by showing that a measure is related to other measures as specified in a theory.

Reliability

Reliability means that a measurement yields consistent scores (so scores change only when the phenomenon changes). If a measure is *reliable*, it is affected less by random error, or chance variation, than if it is unreliable. Reliability is a prerequisite for measurement validity: We cannot really measure a phenomenon if the measure we are using gives inconsistent results. Let's say, for example, that you would like to know your weight and have decided on two different measures: the scales in the bathroom and your mother's estimate. Clearly, the scales are more reliable, in the sense that they will show pretty much the same thing from one day to the next unless your weight actually changes. But your mother, bless her, may say, "You're so skinny!" on Sunday, but on Monday, when she's not happy, she may say, "You look terrible! Have you gained weight?" Her estimates may bounce around quite a bit. The bathroom scales are not so fickle; they are *reliable.*

Audio Link

Listen to what reliability is and how it affects social research.

This doesn't mean that the scales are *valid*—in fact, if they are spring-operated and old, they may be off by quite a few pounds. But they will be off by the same amount every day—hence not being valid but *reliable* nonetheless.

Establishing reliability of a measure is much more straightforward than establishing validity. Essentially, you will be comparing the measure with itself, in various ways. For example, a test of your knowledge of research methods would be unreliable if every time you took it, you received a different score, even though your knowledge of research methods had not changed in the interim. This is **test–retest reliability**. The test would have **interitem reliability** if doing well on some questions (items) matched up with doing well on others. When the wording of questions is altered slightly, your overall grade should still stay roughly the same (**alternate-forms reliability**). If you make an A on the first half of the test, you shouldn't get an F on the second half (**split-halves reliability**). Finally, whether your professor, or your TA, or another expert in the field evaluates your test shouldn't affect your grade (**interobserver reliability)**.

Reliability: A measurement procedure yields consistent scores when the phenomenon being measured is not changing.

Test–retest reliability: A measurement showing that measures of a phenomenon at two points in time are highly correlated, if the phenomenon has not changed or has changed only as much as the phenomenon itself.

Interitem reliability (internal consistency): An approach that calculates reliability based on the correlation between multiple items used to measure a single concept.

Alternate-forms reliability: A procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

Split-halves reliability: Reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.

Interobserver reliability: When similar measurements are obtained by different observers rating the same persons, events, or places.

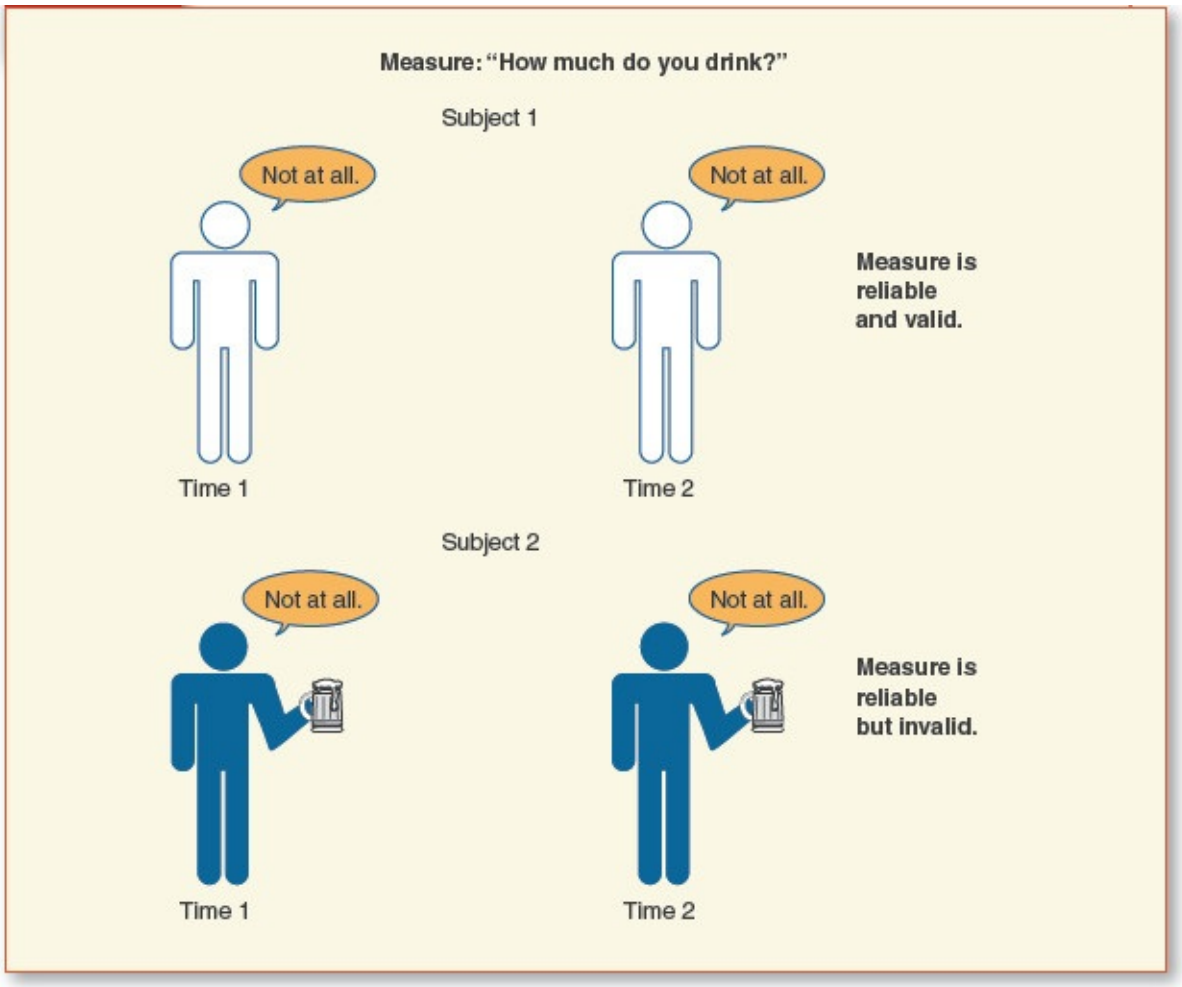
Can We Achieve Both Reliability and Validity?

The reliability and validity of measures in any study must be tested after the fact to assess the quality of the information obtained. But then, if it turns out that a measure cannot be considered reliable and valid, little can be done to save the study. Hence, it is supremely important to select in the first place measures that are likely to be both reliable and valid. The Dow Jones Industrials Index is a perfectly *reliable* measure of the state of the U.S. economy—any two observers of it will see the same numbers—but its validity is shaky: There's more to the economy than the rise and fall of stock prices. In contrast, a good therapist's interview of a married couple may produce a *valid* understanding of their relationship, but such interviews are often not reliable because another interviewer could easily reach different conclusions.

Finding measures that are both reliable and valid can be challenging. Don't just choose the first measure you find or can think of. Consider the different strengths of different measures and their appropriateness to your study. Conduct a pretest in which you use the measure with a small sample and check its reliability. Provide careful training to ensure a consistent approach if interviewers or observers will administer the measures. In most cases, however, the best strategy is to use measures that have been used before and whose reliability and validity have been established in other contexts. But even the selection of "tried and true" measures does not absolve researchers from the responsibility of testing the reliability and validity of the measure in their own studies.

Remember that a reliable measure is not necessarily a valid measure, as Exhibit 4.6 illustrates. The discrepancy shown is a common flaw of self-report measures of substance abuse. People's answers to the questions are consistent (reliable), but they are consistently misleading (not valid): A number of respondents will not admit to drinking, even though they drink a lot. Most respondents answer the multiple questions in self-report indexes of substance abuse in a consistent way, so the indexes are reliable. As a result, some indexes based on self-report are reliable but invalid. Such indexes are not useful and should be improved or discarded.

Exhibit 4.6 The Difference Between Reliability and Validity: Drinking Behavior



Research|Social Impact Link

Read more about reliability and validity in the field of forensics.

Conclusion

Remember always that measurement validity is a necessary foundation for social research. Gathering data without careful conceptualization or conscientious efforts to operationalize key concepts often is a wasted effort.

The difficulties of achieving valid measurement vary with the concept being operationalized and the circumstances of the particular study. The examples in this chapter of difficulties in achieving valid measures should sensitize you to the need for caution.

Planning ahead is the key to achieving valid measurement in your own research; careful evaluation is the key to sound decisions about the validity of measures in others' research. Statistical tests can help you determine whether a given measure is valid after data have been collected, but if it appears after the fact that a measure is invalid, little can be done to correct the situation. If you cannot tell how key concepts were operationalized when you read a research report, don't trust the findings. And if a researcher does not indicate the results of tests used to establish the reliability and validity of key measures, remain skeptical.

Key Terms

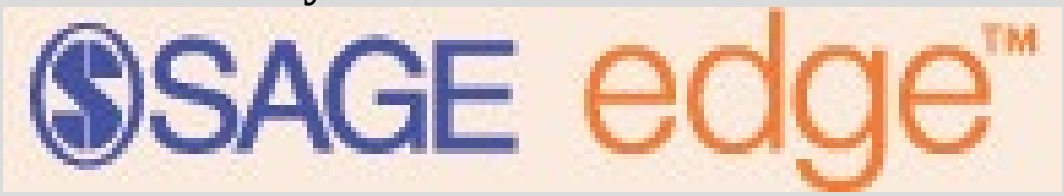
Alternate-forms reliability 83
Closed-ended (fixed-choice) question 72
Concept 66
Conceptualization 67
Constant 68
Construct validity 82
Content analysis 72
Criterion validity 81
Exhaustive 73
Face validity 81
Index 73
Interitem reliability (internal consistency) 83
Interobserver reliability 83
Interval level of measurement 78
Level of measurement 76
Mutually exclusive 73
Nominal level of measurement 76
Open-ended question 73
Operation 69
Operationalization 69
Ordinal level of measurement 78
Ratio level of measurement 78
Reliability 82
Scale 75
Split-halves reliability 83
Test–retest reliability 83
Triangulation 76

Highlights

- Conceptualization plays a critical role in research. In deductive research, conceptualization guides the operationalization of specific variables; in inductive research, it guides efforts to make sense of related observations.
- Concepts may refer to either constant or variable phenomena. Concepts that refer to variable phenomena may be very similar to the actual variables used in a study, or they may be much more abstract.
- Concepts are operationalized in research by one or more indicators, or measures, which may derive from observation, self-report, available records or statistics, books and other written documents, clinical indicators, discarded materials, or some combination.
- Indexes and scales measure a concept by combining answers to several questions and so reducing idiosyncratic variation. Several issues should be explored with every intended index: Does each question actually measure the same concept? Does combining items in an index obscure important relationships between individual questions and other variables? Is the index multidimensional?
- If differential weighting, based on differential information captured by questions, is used in the calculation of index scores, then we say that the questions constitute a scale.
- Level of measurement indicates the type of information obtained about a variable and the type of statistics that can be used to describe its variation. The four levels of measurement can be ordered by complexity of the mathematical operations they permit: nominal (or qualitative), ordinal, interval, and ratio (most complex). The measurement level of a variable is determined by how the variable is operationalized.
- The validity of measures should always be tested. There are three basic approaches: face validation, criterion validation, and construct validation. Criterion validation provides the strongest evidence of measurement validity, but often there is no criterion to use in validating social science measures.
- Measurement reliability is a prerequisite for measurement validity, although reliable measures are not necessarily valid. Reliability can be assessed through a test–retest procedure, an interitem comparison of responses to component measures within an index, a comparison of responses to alternate forms of the test or by randomly selected (“split”) halves of a

sample to the same test, or the consistency of findings among observers.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. What does *trust* mean to you? Identify two examples of “trust in action,” and explain how they represent your concept of trust. Now develop a short definition of *trust* (without checking a dictionary). Compare your definition to those of your classmates and what you find in a dictionary. Can you improve your definition based on some feedback?
2. What questions would you ask to measure the level of trust among students? How about feelings of being “in” or “out” with regard to a group? Write five questions for an index, and suggest response choices for each. How would you validate this measure using a construct validation approach? Can you think of a criterion validation procedure for your measure?
3. If you were given a questionnaire right now that asked you about your use of alcohol and illicit drugs in the past year, would you disclose the details fully? How do you think others would respond? What if the questionnaire was anonymous? What if there was a confidential ID number on the questionnaire so that the researcher could keep track of who responded? What criterion validation procedure would you suggest for assessing measurement validity?

Finding Research

1. What are some of the research questions you could attempt to answer with available statistical data? Visit your library and ask for an introduction to the government documents collection. Inspect the U.S. Census Bureau website (www.census.gov) and find the population figures broken down by city and state. List five questions that you could explore with such data. Identify six variables implied by these research questions that you could operationalize with the available data. What are three factors that might influence variation in these measures other than the phenomenon of interest? (Hint: Consider how the data are collected.)

2. How would you define *alcoholism?* Write a brief definition. Based on this conceptualization, describe a method of measurement that would be valid for a study of alcoholism (as you define it).

   Now go to the American Council for Drug Education, an affiliate of Phoenix House, and read some their facts about alcohol (http://www.phoenixhouse.org/prevention/). Is this information consistent with your definition?

Critiquing Research

1. Shortly before the year 2000 national census of the United States, a heated debate arose in Congress about whether instead of a census—a total headcount—a sample should be used to estimate the number and composition of the U.S. population. As a practical matter, might a sample be more accurate in this case than a census? Why?
2. Develop a plan for evaluating the validity of a measure. Your instructor will give you a copy of a questionnaire actually used in a study. Pick out one question and define the concept that you believe it is intended to measure. Then develop a construct validation strategy involving other measures in the questionnaire that you think should be related to the question of interest—if it measures what you think it measures.
3. The questions in Exhibit 4.7 are selected from a survey of homeless shelter staff (Schutt & Fennell 1992). First, identify the level of measurement for each question. Then rewrite each question so that it measures the same variable but at a different level. For example, you might change a question that measures age at the ratio level, in years, to one that measures age at the ordinal level, in categories. Or you might change a variable measured at the ordinal level to one measured at the ratio level. For the categorical variables, those measured at the nominal level, try to identify at least two underlying quantitative dimensions of variation and write questions to measure variation along these dimensions. For example, you might change a question asking which of several factors the respondent thinks is responsible for homelessness to a series of questions that ask how important each factor is in generating homelessness.
4. What are the advantages and disadvantages of phrasing each question at one level of measurement rather than another? Do you see any limitations on the types of questions for which levels of measurement can be changed?

Exhibit 4.7 Selected Shelter Staff Survey Questions

1. What is your current job title? _____

2. What is your current employment status?

   Paid, full-time _____ 1
   Paid, part-time (less than 30 hours per week) _____ 2

3. When did you start your current position? _____ / _____ / _____
   Month Day Year

4. In the past month, how often did you help guests deal with each of the following types of problems? *(Circle one response on each line.)*

	Very often						Never
Job training/placement	1	2	3	4	5	6	7
Lack of food or bed	1	2	3	4	5	6	7
Drinking problems	1	2	3	4	5	6	7

5. How likely is it that you will leave this shelter within the next year?

   Very likely _____ 1
   Moderately _____ 2
   Not very likely _____ 3
   Not likely at all _____ 4

6. What is the highest grade in school you have completed at this time?

   First through eighth grade _____ 1
   Some high school _____ 2
   High school diploma _____ 3
   Some college _____ 4
   College degree _____ 5
   Some graduate work _____ 6
   Graduate degree _____ 7

7. Are you a veteran?

   Yes _____ 1
   No _____ 2

Source: Based on Schutt, Russell K. 1988. Working with the homeless: The backgrounds, activities and beliefs of shelter staff. Boston: University of Massachusetts. Unpublished report: 7–10, 15, 16. Results reported in Schutt, Russell K., and M. L. Fennell. 1992. Shelter staff satisfaction with services, the service network, and their jobs. Current Research on Occupations and Professions 7: 177–200.

Doing Research

1. Some people have said in discussions of international politics that “democratic governments don’t start wars.” How could you test this hypothesis? Clearly state how you would operationalize (1) *democratic* and (2) *start.*

2. Now it’s time to try your hand at operationalization with survey-based measures. Formulate a few fixed-choice questions to measure variables pertaining to the concepts you researched for Exercise 1 under “Discussing Research.” Arrange to interview one or two other students with the questions you have developed. Ask one fixed-choice question at a time, record your interviewee’s answer, and then probe for additional comments and clarifications. Your goal is to discover what respondents take to be the meaning of the concept you used in the question and what additional issues shape their response to it.

   When you have finished the interviews, analyze your experience: Did the interviewees interpret the fixed-choice questions and response choices as you intended? Did you learn more about the concepts you were working on? Should your conceptual definition be refined? Should the questions be rewritten, or would more fixed-choice questions be necessary to capture adequately the variation among respondents?
3. Now try index construction. You might begin with some of the questions you wrote for Exercise 2. Write four or five fixed-choice questions that each measure the same concept. (For instance, you could ask questions to determine whether someone is alienated.) Write each question so it has the same response choices (a matrix design). Now conduct a literature search to identify an index that another researcher used to measure your concept or a similar concept. Compare your index to the published index. Which seems preferable to you? Why?
4. List three attitudinal variables.
    1. Write a conceptual definition for each variable. Whenever possible, this definition should come from the existing literature—either a book you have read for a course or the research literature that you have searched. Ask two class members for feedback on your definitions.
    2. Develop measurement procedures for each variable: Two measures should be single questions, and one should be an index used in prior research (search the Internet and the journal literature in Sociological Abstracts or Psychological Abstracts). Ask classmates to answer these questions and give you feedback on their clarity.
    3. Propose tests of reliability and validity for the measures.
5. Exercise your cleverness on this question: For each of the following, suggest two unobtrusive measures that might help you discover (a) how much of the required reading for this course students actually complete, (b) where are the popular spots to sit in a local park, and (c) which major U.S. cities have the highest local taxes.

Ethics Questions

1. The ethical guidelines for social research require that subjects give their “informed consent” before participating in an interview. How “informed” do you think subjects have to be?

   If you are interviewing people to learn about substance abuse and its impact on other aspects of health, is it okay just to tell respondents in advance that you are conducting a study of health issues? What if you plan to inquire about victimization experiences? Explain your reasoning.
2. Both some Homeland Security practices and inadvertent releases of web searching records have raised new concerns about the use of unobtrusive measures of behavior and attitudes. If all identifying information is removed, do you think social scientists should be able to study the extent of prostitution in different cities by analyzing police records? How about how much alcohol different types of people use by linking credit card records to store purchases?

Video Interview Questions

Listen to the researcher interview for Chapter 4 at edge.sagepub.com/chamblissmssw5e.

1. What problems does Dana Hunt identify with questions designed to measure frequency of substance abuse and aggressive feelings?
2. What could be done to overcome these problems?

Chapter 5 Sampling and Generalizability

©iStockphoto.com/busypix

Learning Objectives

1. Distinguish the two foci of sampling theory.
2. Identify the circumstances that make sampling unnecessary and the reason they're rare.
3. Identify the relation between the elements, the sampling units, the sample, the sampling frame, and the target population.
4. Define the concept of representative sample and explain how it contrasts with the concept of bias.
5. Define and distinguish probability and nonprobability sampling.
6. Define the major types of probability sampling method and indicate when each is preferred.
7. Explain when nonprobability sampling methods may be preferred.

An old history professor was renowned for his ability, at semester's end, to finish grading large piles of student papers (many of them undistinguished) in a matter of a few short hours. When asked by a younger colleague how he accomplished this feat, the codger replied with a snort, "You don't have to eat the whole tub of butter to know if it's rancid." Harsh, but true.

That is the essence of sampling: A small portion, carefully chosen, can reveal the quality of a much larger whole. A survey of 1,400 Americans telephoned one Saturday afternoon can tell us quite accurately how 40 million will vote for president on the following Tuesday morning. A quick check of reports from a few selected banks can tell the Federal Reserve how strong inflation is. And when you go to the health clinic with a possible case of mononucleosis and a blood test is done, the phlebotomist needn't take all of your blood to see if you have too many atypical lymphocytes. Sampling techniques tell us how to select cases that can lead to valid generalizations about a **population**, or the entire group you want to learn about. In this chapter, we define the key components of sampling strategy and then present the types of sampling one may use in a research study along with the strengths and weaknesses of each.

Population: The entire set of individuals or other entities to which study findings are to be generalized.

How Do We Prepare to Sample?

Define Sample Components and the Population

To understand how sampling works, you'll first need a few useful definitions. A **sample** is a subset of the population that we want to learn about. For instance, suppose the human resources (HR) offices at a large retail clothing chain wants to understand the career aspirations of their employees. The population would be all current employees of the company. The sample could be, say, 200 individuals whom HR will select to interview. The individual members of this sample are called **elements**—that is, the specific people selected. These are the cases that we actually study. To select these elements, we often rely on some list of all elements in the population—**a sampling frame**. In our example, this would be a list of all current employees. In some cases, a sampling frame may be quite difficult to produce: all homeless people in Chicago, all drug users at your universities, or all professional comedians in San Francisco.

A sample can only represent the population from which it was drawn. So if we sample students in one high school, the population for our study is the student body of that school, not all high school students in the nation. Some populations, such as frequent moviegoers, are not identified by a simple criterion, such as a geographic boundary or an organizational membership. Clear definition of such a population is difficult but quite necessary. Anyone should be able to determine just what population was actually studied, so we would have to define clearly the concept of *frequent moviegoers* and specify how we determined their status.

Often researchers make fundamental sampling mistakes even before they start examining their data, for instance, by selecting the wrong sampling frame—one that does not adequately represent the population. Perhaps the most common version of this error is called *sampling on the dependent variable*, in which cases are chosen not to represent the population but because they represent a (usually) interesting outcome—that is, only one value of the dependent variable. Even the best social scientists sometimes fall into this trap. In their fascinating and important book *Rampage: The Social Roots of School Shootings*, Katherine S. Newman and her coauthors studied in detail the case histories of 27 different teenagers who had gone into their schools and killed (mostly random) fellow students—the Columbine attack of April 20, 1999, may be the most famous case, where the shooters killed 13 and wounded 21 others, then killed themselves (Newman et al. 2004). You may be familiar with the 2012 Sandy Hook

Elementary School shootings, when 20-year-old Adam Lanza fatally shot 20 young children and 6 adult teachers in Newtown, Connecticut. Based on their study of school shooters, Newman and colleagues concluded that there were five "necessary but not sufficient" factors in school shootings: (1) a self-perception of shooters as socially marginal, (2) psychosocial problems, (3) cultured scripts linking masculinity and violence, (4) failure of surveillance systems (so troubled kids are "under the radar"), and (5) availability of guns. Virtually all school shooters fit this description; they have all these characteristics. *Rampage* is a valuable piece of serious exploratory social science.

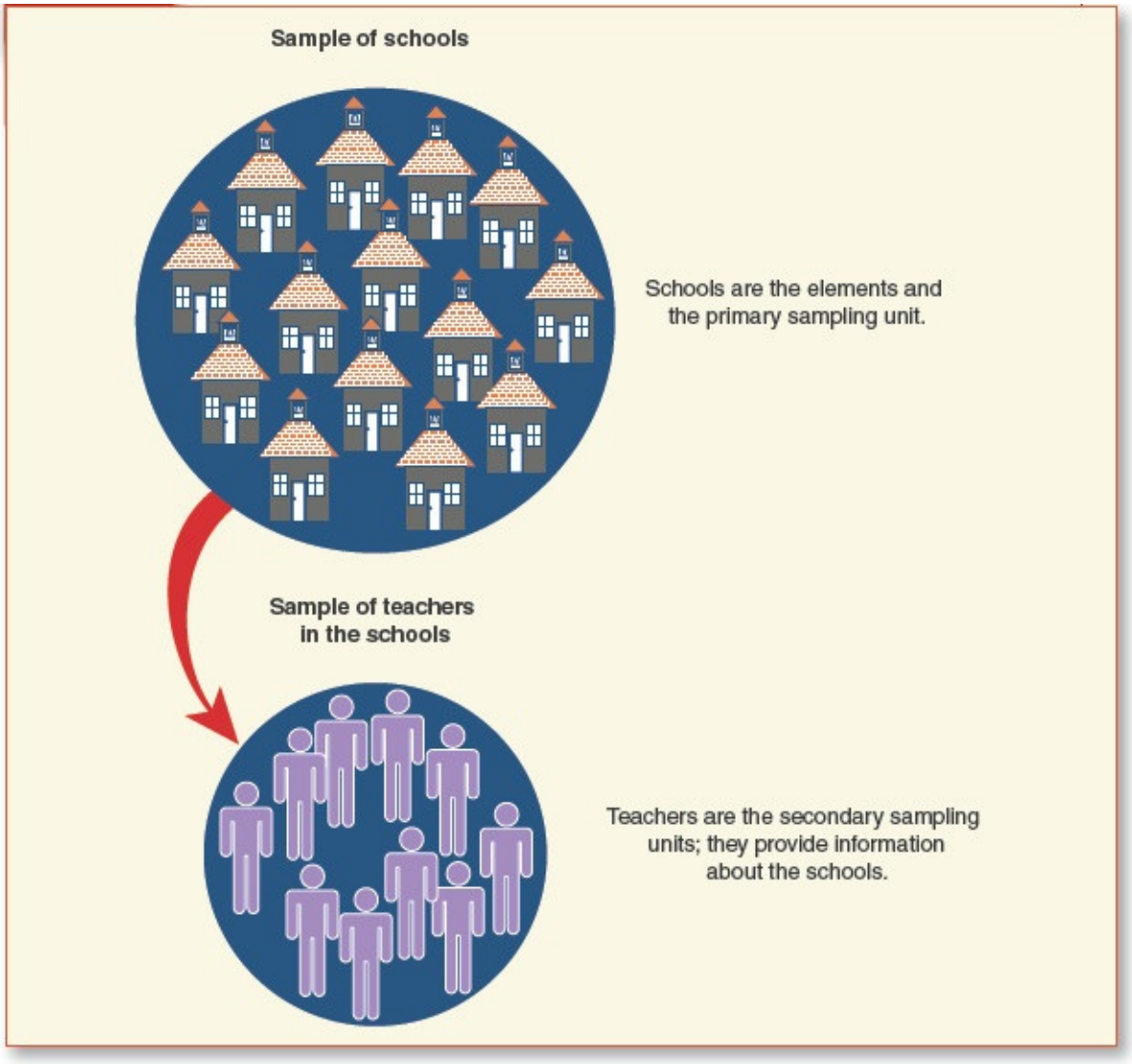
But this model still does not explain shootings, or even tell us much about who will commit them. The fact is, all of the shooters were also boys, they were all teenagers, and they all attended high school. Were these also important factors in explaining their participation in the school shootings? And were there other students who perceived themselves as socially marginal, or who had psychosocial problems, and so on? Why didn't these other students turn into school shooters? The problem, in other words, is that Newman and her colleagues (2004) only looked at shooters, instead of comparing shooters with nonshooters to see what made the difference. Their sampling frame (a list of school shooters) allowed them to generalize to other school shooters, but not to tell you how shooters differ from other teenagers.

Video Link

Watch a guide to the types of sampling and how they relate to the population.

Sometimes our sources of information are not actually the elements in our study. For example, for a survey about educational practices, a researcher might first sample schools and then, within sampled schools, interview a sample of teachers. The schools and the teachers are both termed **sampling units** because the researcher sampled from both (Levy & Lemeshow 1999: 22). The schools are selected in the first stage of the sample, so they are the *primary sampling units* (and in this case, the elements in the study). The teachers are *secondary sampling units* (but they are not elements because they are used to provide information about the entire school) (Exhibit 5.1).

Exhibit 5.1 Sample Components in a Two-Stage Study


Sample of schools
Schools are the elements and the primary sampling unit.
Sample of teachers in the schools
Teachers are the secondary sampling units; they provide information about the schools.

Sample: A subset of a population used to study the population as a whole.

Elements: The individual members of the population whose characteristics are to be measured.

Sampling frame: A list of all elements or other units containing the elements in a population.

Sampling units: Units listed at each stage of a multistage sampling design.

Evaluate Generalizability

Once we have defined clearly the population from which we will sample, we need to determine the scope of the generalizations we will seek to make from our sample. Do you recall the two different meanings of *generalizability* from Chapter 1?

Journal Link

Read an article that uses data from multiple countries to reach generalizable results.

- *Can the findings from a sample of the population be generalized to the population from which the sample was selected?* This issue was defined in Chapter 1. Again, when the Gallup polls ask some Americans for their political opinions, can those answers be generalized to the U.S. population? Probably so. But if Gallup's sampling was haphazard—say, if the pollsters just talked to some people in the office—they probably couldn't make the same accurate generalizations.
- *Can the findings from a study of one population be generalized to another, somewhat different population?* Are residents of three impoverished communities in the city of Enschede, the Netherlands, similar to those in other communities? In other cities? In other nations? The problem here was defined in Chapter 1 as *cross-population generalizability.* For example, many psychology studies are run using (easily available) college students as subjects. Because such research is often on tasks that require no advanced education, such as memorizing lists of nonsense syllables or spotting patterns in an array of dots, college students may in this respect be like most other human beings, so the generalization seems legitimate. But when psychoanalyst Sigmund Freud talked with a very narrow sample of Viennese housewives in 1900, could his findings be accurately generalized (as he attempted) to the entire human race? Probably not.

Journal Link

Read an article and assess generalizability based on its sample of 19–24 year olds.

This chapter focuses attention primarily on the problem of sample generalizability: Can findings from a sample be generalized to the population from which the sample was drawn? This is really the most basic question to ask about a sample, and social research methods provide many tools with which to address it.

But researchers often project their theories onto groups or populations much larger than, or simply different from, those they have actually studied. The population to which generalizations are made in this way can be termed the **target population**—a set of elements larger than or different from the population that was sampled and to which the researcher would like to generalize any study findings. Because the validity of cross-population generalizations cannot be tested empirically, except by conducting more research in other settings, we will not focus much attention on this problem here.

Target population: A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.

Assess the Diversity of the Population

Sampling is unnecessary if all the units in the population are identical. The blood in one person is constantly being mixed and stirred, so it's very homogeneous—any pint is the same as any other. Nuclear physicists don't need a representative sample of all atomic particles to learn about basic atomic processes because in crucial respects all such particles are alike.

Video Link

Watch a clip on how the US Census Bureau uses their research to inform social policy.

What about people? Certainly all people are not identical, but if we are studying fundamental physical or psychological processes that are the same among all people, sampling is not needed to achieve generalizable findings. Psychologists and social psychologists often conduct experiments on college students to learn about such processes (basic cognitive functioning, for instance). But we must always bear in mind that we don't really know how generalizable our findings are to populations that we haven't actually studied.

So, we usually must study the larger population in which we are interested if we want to make generalizations about it. For this purpose, we must obtain a **representative sample** of the population to which generalizations are sought (Exhibit 5.2).

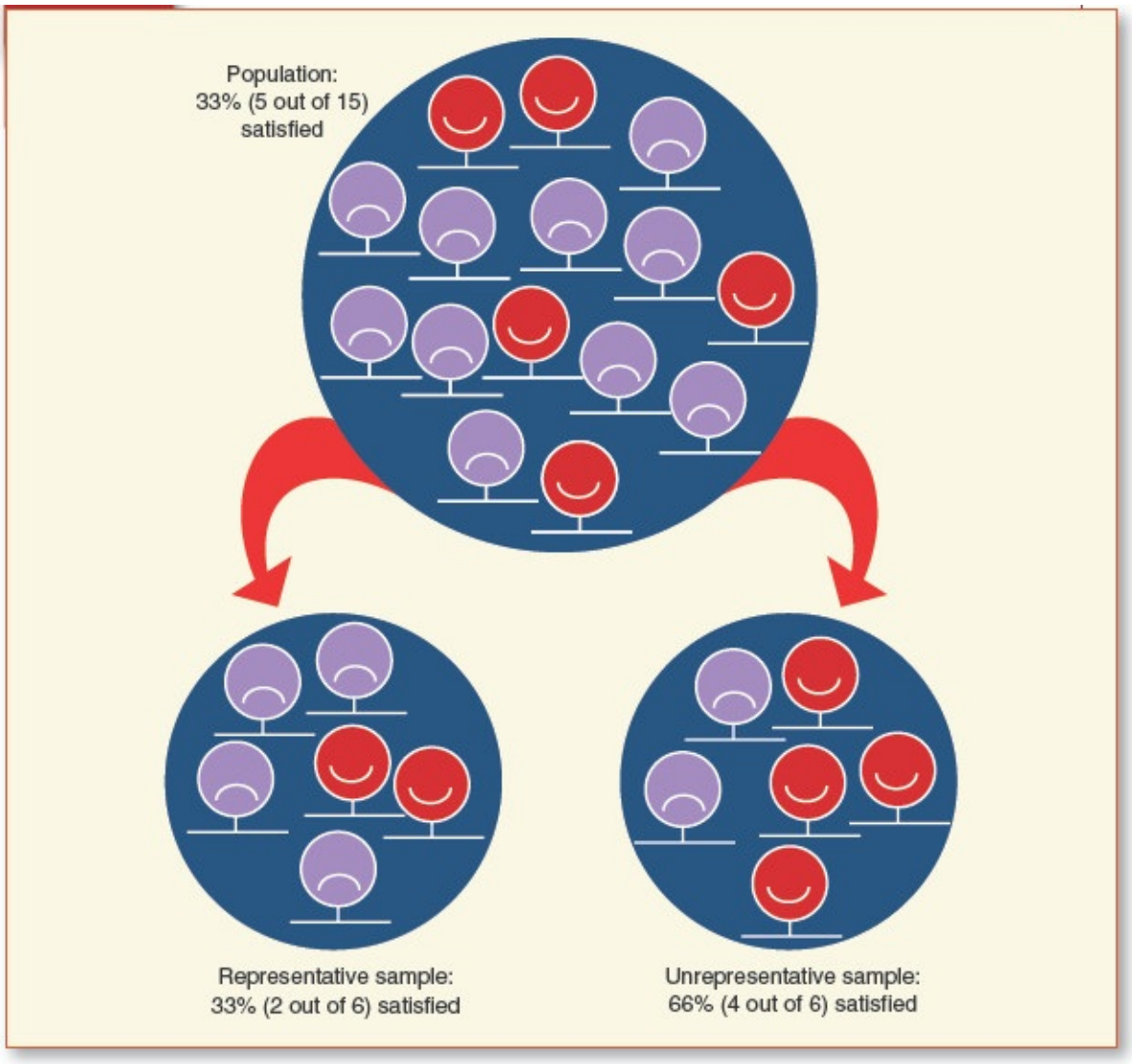
> **Representative sample**: A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Consider a Census

In some circumstances, it may be feasible to establish generalizability by simply conducting a **census**—studying an entire population—rather than drawing a sample. This is what the federal government tries to do every 10 years with the U.S. Census. Censuses could also include, for instance, studies of all the employees in a small business, studies comparing all 50 states, or studies of all the museums in some region.

Social scientists don’t often attempt to collect data from all the members of some large population because doing so would be too expensive and time consuming. The 2010 U.S. Census, for example, is estimated to have cost around $15 billion, or about $48 per person counted. But fortunately, a well-designed sampling strategy can result in a representative sample of the same population at far less cost.

Exhibit 5.2 Representative and Unrepresentative Samples



Census: Research in which information is obtained through responses from or information about all available members of an entire population.

Research That Matters

Homeless populations are especially difficult to sample in representative ways, so little is known about how many homeless young adults are employed and what distinguishes them from the unemployed. Kristin Ferguson and her colleagues Kimberly Bender, Sanna Thompson, Elaine Maccio, and David Pollio (2012: 389–390) decided to interview homeless young adults in five U.S. cities in different regions of the country. The researchers secured the cooperation of multiservice, nonprofit organizations that provide comprehensive services to homeless youth and then, accompanied by agency staff, approached youth in these agencies and on the streets.

One of their findings was that young adults in different cities varied in their employment status and sources of income. For example, homeless young adults in Los Angeles were more likely to be employed, and young adults in Austin, Texas, were significantly more likely to receive their income from panhandling (Ferguson et al. 2012: 400). Drawing a representative sample is often very difficult, particularly in studies of hard-to-reach groups such as homeless youth.

Source: Adapted from Ferguson, Kristin M., Kimberly Bender, Sanna J. Thompson, Elaine M. Maccio, and David Pollio. 2012. Employment status and income generation among homeless young adults: Results from a five-city, mixed-methods study. *Youth & Society* 44: 385–407.

What Sampling Method Should We Use?

Certain features of samples make them more or less likely to represent the population from which they are selected; the more representative the sample, the better. The crucial distinction about samples is whether they are based on a probability or a nonprobability sampling method. **Probability sampling methods** allow us to know in advance how likely it is that any element of a population will be selected. Sampling methods that do not let us know in advance the likelihood of selecting each element are termed **nonprobability sampling methods**.

Probability sampling methods rely on a random, or chance, selection procedure, which is in principle the same as flipping a coin to decide which of two people "wins" and which one "loses." Heads and tails are equally likely to turn up in a coin toss, so both persons have an equal chance to win. That chance, their **probability of selection**, is 1 out of 2, or 0.5.

There is a natural tendency to confuse the scientific concept of **random sampling**, in which cases are selected only on the basis of chance, with *haphazard* sampling. On first impression, "leaving things up to chance" seems to imply not exerting any control over the sampling method. But to achieve true randomness, the researcher must proceed very methodically, following careful procedures. With random sampling, every element (every person, in many studies) has the same chance of being selected, so that the sample will more accurately represent the entire population.

Two common problems can bias even what appear to be random samples:

1. If the sampling frame is incomplete, a random sample from that list will not really be a random sample of the population. You should always consider the adequacy of the sampling frame. Even for a fairly small population such as a university's student body, the registrar's list is likely to be at least somewhat out-of-date at any given time—and the missing students are probably different from those in the list.
2. Nonresponse is a major hazard, especially in survey research, because nonrespondents are likely to differ systematically from those who take the time to. If the response rate is low (say, below 65%), then, you won't really be getting the random sample that you originally chose, and you should not

assume that findings from even a good random sample will be generalizable to the population.

Probability Sampling Methods

Introduced earlier, probability sampling methods are those in which the probability of selection is known and is not zero (so there is some chance of selecting each element). These methods randomly select elements and therefore have no systematic **bias**; nothing but chance determines which elements are included in the sample. When the goal is to generalize to a larger population, then, probability samples are much useful than nonprobability (biased) samples are.

However, even a randomly selected sample will always have some degree of sampling error—some deviation from the characteristics of the population. If you randomly choose 10 Americans (say, by a lottery that includes everyone) to learn what Americans generally think about abortion, they may not be very typical—you might, just by chance, have picked up 8 women and only 2 men, for instance. It would help to get more people, at least until the sample "smooths out" the proportions of such groups. Your job also would be easier, of course, if everyone had similar opinions. Formally stated, both the size of the sample and the homogeneity (sameness) of the population affect the degree of error due to chance. It helps, to a point, to have more people, and it definitely helps if everyone is the same, but that's not usually the case! Interestingly, the *proportion* of the total population represented by the sample (10%, 20%, etc.) does not affect its representativeness, unless that proportion is very large; the raw number of cases in the sample is what is important. To represent Americans, for instance, once you have more than 1,000 or so people, adding still more to your sample doesn't really help very much—and the information gained from each new person diminishes the more you add.

To elaborate,

- *The larger the sample, the more confidence we can have in the sample's representativeness.* If we randomly pick 5 people to represent the entire population of our city, our sample is unlikely to be very representative of the entire population in terms of age, gender, race, attitudes, and so on. But if we randomly pick 100 people, the odds of having a representative sample are much better; with a random sample of 1,000, the odds become very good indeed.

- *The more homogeneous the population, the more confidence we can have in the representativeness of a sample of any particular size*. That's why blood testing works—blood is homogeneous in any one person's body. Or, let's say we plan to draw samples of 50 people from each of two communities to estimate the mean family income. One community is very diverse, with family incomes varying from $12,000 to $85,000. In the other, more homogeneous community, family incomes are concentrated in a narrow range, from $41,000 to $64,000. The estimated mean family income based on the sample from the homogeneous community is more likely to be representative than is the estimate based on the sample from the more heterogeneous community. With less variation to represent, fewer cases are needed to represent the homogeneous community.

Encyclopedia Link

Read an overview of probability sampling.

Again, the fraction of the total population contained in a sample does not affect the sample's representativeness, unless that fraction is really large. This isn't obvious, but it is mathematically true. The raw number of cases—getting those first few hundred, up to 1,000 or so—matters more than the final proportion of the population. The larger size of the sample is what makes it more representative, not the proportion of the whole that the sample represents.

Polls to predict presidential election outcomes illustrate both the value of random sampling and the problems that it cannot overcome. In most presidential elections, pollsters have predicted fairly accurately the outcomes of the actual votes by using random sampling and, these days, phone interviewing to learn for whom likely voters intend to vote. Exhibit 5.3 shows how accurate these sample-based predictions have been in the last 14 contests. The exceptions were the 1980 and 1992 elections, when third-party candidates had a surprising effect. Otherwise, the small discrepancies between the votes predicted through random sampling and the actual votes can be attributed to random error.

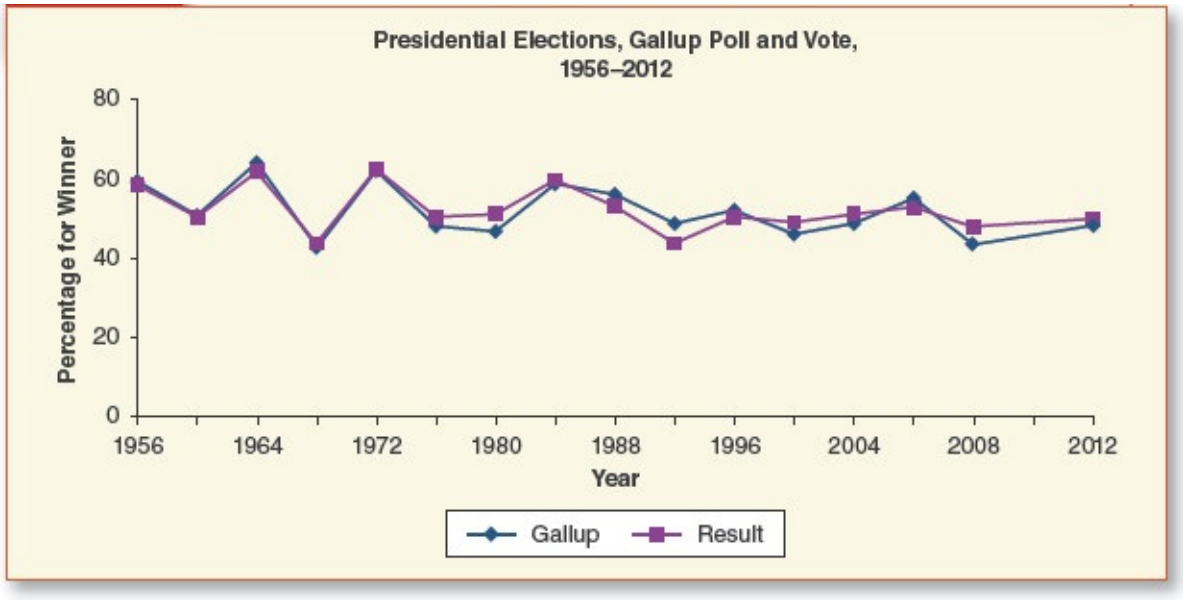
Journal Link

Read about survey data collected through probability sampling.

The Gallup poll, one of the oldest and most respected, did not do quite as well in predicting the results of the 2008 and 2012 presidential elections as some other polling organizations did. The final 2008 Gallup prediction was that Barack Obama would win with 55% to John McCain's 44% (Gallup 2011), but Obama actually won 53% to 46%. In 2012, Gallup predicted that Mitt Romney would win by 1%, but the reverse actually occurred. Nonetheless, contemporary pollsters do come remarkably close to the final results. But even well-run election polls have produced some major errors in prediction. In 1948, pollsters mistakenly predicted that Thomas E. Dewey would beat Harry S. Truman, based on the random sampling method that George Gallup had used successfully since 1934. The problem? Pollsters stopped collecting data several weeks before the election, and in those weeks, many people changed their minds (Kenney 1987). So just as in the 2000 election, underrepresenting shifts in voter sentiment just before the election systematically biased the sample.

Every method of sampling has its uses and its disadvantages; depending on the purpose of your research, you'll need to choose the one that works best. Probability-based sampling is certainly preferable most of the time, but isn't always feasible. We'll examine four probability and four nonprobability sampling techniques here, pointing out the pros and cons of each.

Exhibit 5.3 Presidential Election Outcomes: Predicted and Actual



Source: Gallup Organization. 2011. Election polls—Accuracy record in presidential elections. http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print (accessed March 17, 2011);

The four most common types of probability (random) sample are (1) simple random sampling, (2) systematic random sampling, (3) cluster sampling, and (4) stratified random sampling.

Simple Random Sampling

Simple random sampling, the scientifically most "pure" approach, identifies cases strictly on the basis of chance. It will most accurately represent the population you are studying. Flipping a coin or rolling a die can be used to identify cases strictly on the basis of chance, but these procedures are not very efficient tools for drawing a sample from large sampling frames. A **random number table** simplifies the process considerably. The researcher numbers all the elements in the sampling frame and then uses a systematic procedure for picking corresponding numbers from the random number table. (Exercise 1 under "Doing Research" at the end of this chapter explains the process step-by-step.) Alternatively, a researcher may use a lottery procedure. Each case number is written on a small card, and then the cards are mixed up and the sample selected from the cards. A computer program can also easily generate a random sample of any size.

Phone surveys often use a technique called **random digit dialing (RDD)** to draw a random sample. A machine dials random numbers within the phone prefixes corresponding to the area in which the survey is to be conducted. Random digit dialing is particularly useful when a sampling frame (list of elements) is unavailable because the dialing machine can just skip ahead if a phone number is not in service.

In a true simple random sample, the probability of selection is equal for each element. If a sample of 500 is selected from a population of 17,000 (that is, a sampling frame of 17,000), then the probability of selection for each element is 500/17,000, or 0.03. Every element has an equal chance of being selected, just like the odds in a toss of a coin (1/2) or a roll of a die (1/6). Thus, simple random sampling is an *equal probability of selection method (EPSEM).*

Probability sampling method: A sampling method that relies on a random, or chance, selection method so that the probability of selection of population elements is known.

Nonprobability sampling methods: Sampling methods in which the probability of selection of population elements is unknown.

Probability of selection: The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all the elements of a population, the probability that any particular element will be selected is 1.0. If half the elements in the population are sampled on the basis of chance (say, by tossing a coin), the probability of selection for each element is one half, or 0.5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

Random sampling: A method of sampling that relies on a random, or chance, selection method so that every element of the sampling frame has a known probability of being selected.

Bias: Sampling bias occurs when some population characteristics are over- or underrepresented in the sample because of particular features of the method of selecting the sample.

Research in the News

What are Best Practices for Sampling Vulnerable Populations?

In the 1950s, Perry Hudson studied the effectiveness of early prostate screening in reducing cancer. He sampled 1,200 alcoholic homeless men from the flophouses of Lower Manhattan. His research was funded and supported by the National Institutes of Health, but he did not properly inform participants of the risks associated with prostate screening. Many men who participated endured a painful prostate biopsy and no medical follow-up if screened positive for cancer. Robert Aronowitz, a medical historian, looks back at this ethical tragedy as "a convenient population" used in the name of science.

1. Since then, research standards have been changed to protect vulnerable populations. In what types of circumstances do you think that it is ethical to draw samples for research from prisoners, patients, students, and other "captive" populations that are convenient to study?
2. Should samples of large populations exclude persons who suffer from mental illness, addiction, extreme poverty, limited literacy, or other conditions that might make them less likely to make a fully informed decision about participation in research? Are there any risks with such exclusions?

News source: Kolata, Gina. 2013. Decades later, condemnation for a skid row cancer study. *New York Times,* October 18: A1.

Simple random sampling: A method of sampling in which every sample element is selected purely on the basis of chance through a random process.

Random number table: A table containing lists of numbers that are ordered solely on the basis of chance; it is used for drawing a random sample.

Random digit dialing (RDD): The random dialing, by a machine, of numbers within designated phone prefixes, which creates a random sample for phone surveys.

Systematic random sampling: A method of sampling in which sample elements are selected from a list or from sequential files, with every *n*th element being selected after the first element is selected randomly.

Periodicity: A sequence of elements (in a list to be sampled) that varies in some regular, periodic pattern.

Sampling interval: The number of cases between one sampled case and another in a systematic random sample.

Cluster sampling: Sampling in which elements are selected in two or more stages, with the first

stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Systematic Random Sampling

Systematic random sampling is an easy-to-use, efficient variant of simple random sampling. In this method, the first element is selected randomly from a list or from sequential files, and then every *n*th element is selected—for instance, every 7th name on an alphabetical list. This is a convenient method for drawing a random sample when the population elements are arranged sequentially. It is particularly efficient when the elements are not written down (that is, there is no written sampling frame) but instead are represented physically, say, by folders in filing cabinets.

In *almost* all sampling situations, systematic random sampling yields what is essentially a simple random sample; though not as mathematically pure, in practice it works essentially just as well. The exception is a situation in which the sequence of elements is characterized by **periodicity**—that is, the sequence varies in some regular, periodic pattern. For example, in a new housing development with the same number of houses on each block (eight, for example), houses may be listed by block, starting with the house in the northwest corner of each block and continuing clockwise. If the **sampling interval** is 8, the same as the periodic pattern, all the cases selected will be in the same position (Exhibit 5.4). Those houses may well be unusual—corner locations are typically more expensive, for instance. But usually, periodicity and the sampling interval are rarely the same, so this isn't a problem.

Research|Social Impact Link

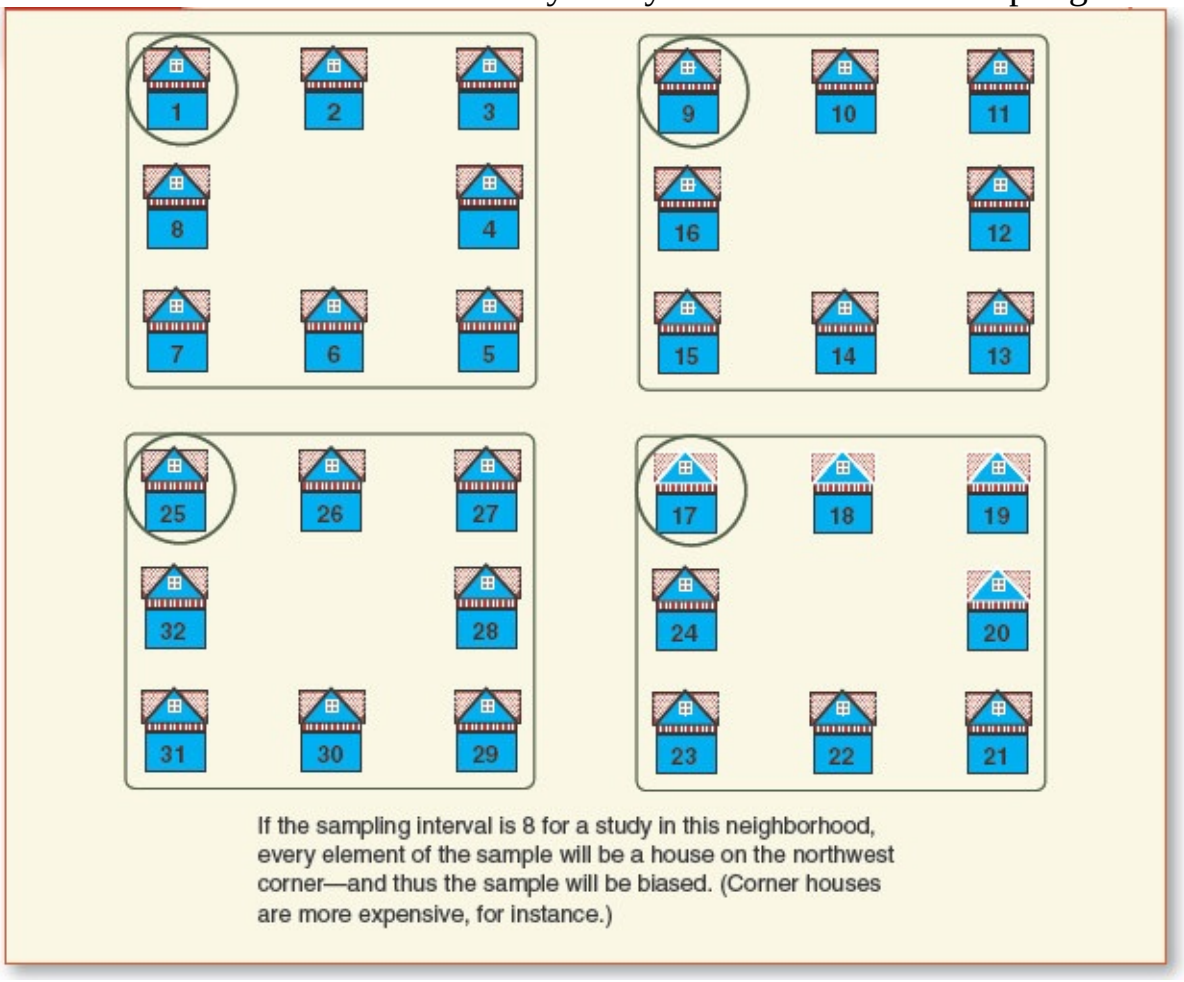
Read about how cluster sampling methods were used in the Iraq War.

Cluster Sampling

Cluster sampling is useful when a sampling frame—a definite list—of elements is not available, as often is the case for large populations spread out across a wide geographic area or among many different organizations. We don't have a good list of all the Catholics in America, all the businesspeople in Arizona, or all

the waiters in New York. A **cluster** is a naturally occurring, mixed aggregate of elements of the population, with each element (person, for instance) appearing in one and only one cluster. Schools could serve as clusters for sampling students, city blocks could serve as clusters for sampling residents, counties could serve as clusters for sampling the general population, and restaurants could serve as clusters for sampling waiters.

Exhibit 5.4 The Effect of Periodicity on Systematic Random Sampling



Source: Gallup Organization. 2011. Election polls—Accuracy record in presidential elections. http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print (accessed March 17, 2011)

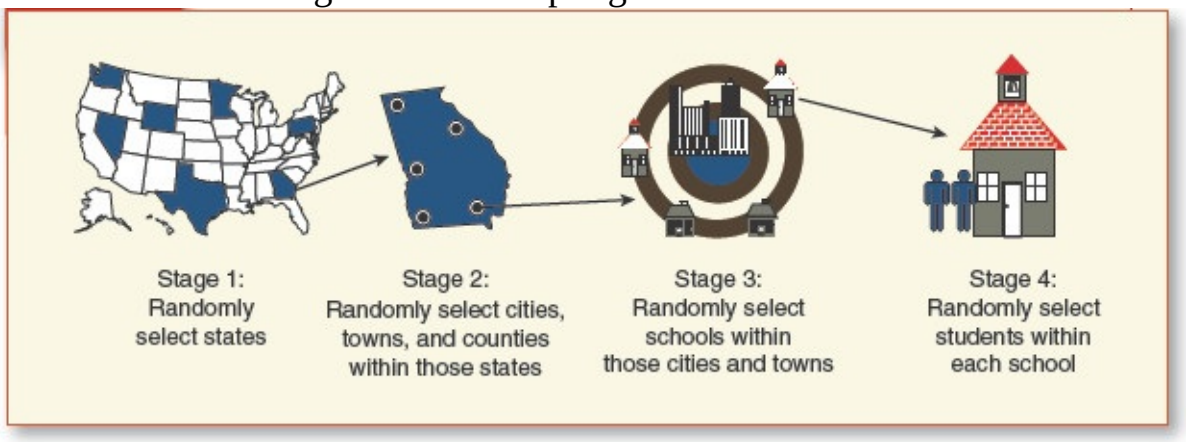
Cluster sampling is at least a two-stage procedure. First, the researcher draws a

random sample of clusters. (A list of clusters should be much easier to obtain than a list of all the individuals in each cluster in the population.) Next, the researcher draws a random sample of elements within each selected cluster. Because only a fraction of the total clusters is involved, obtaining the sampling frame at this stage should be much easier.

Cluster samples often involve multiple stages, with clusters within clusters, as when a national study of middle school students might involve first sampling states, then counties, then schools, and finally students within each selected school (Exhibit 5.5).

How many clusters and how many individuals within clusters should be selected? As a general rule, the more clusters you select, with the fewest individuals in each, the more representative your sampling will be. Unfortunately, this strategy also maximizes the cost of the sample. The more clusters selected, the higher the travel costs. Remember, too, that the more internally homogeneous the clusters, the fewer cases needed per cluster. Homogeneity within a cluster is good.

Exhibit 5.5 Multistage Cluster Sampling



Source: Gallup Organization. 2011. Election polls—Accuracy record in presidential elections. http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print (accessed March 17, 2011)

Audio Link

Listen to how sampling methods mattered in 2012 presidential election polls.

Cluster sampling is a very popular method among survey researchers, but it has one general drawback: Sampling error is greater in a cluster sample than in a simple random sample because there are two steps involving random selection rather than just one. This sampling error increases as the number of clusters decreases, and the sampling error decreases as the homogeneity of cases per cluster increases. This is another way of restating the preceding points: It's better to include as many clusters as possible in a sample, and it's more likely that a cluster sample will be representative of the population if cases are relatively similar within clusters.

Cluster: A naturally occurring, mixed aggregate of elements of the population.

Stratified Random Sampling

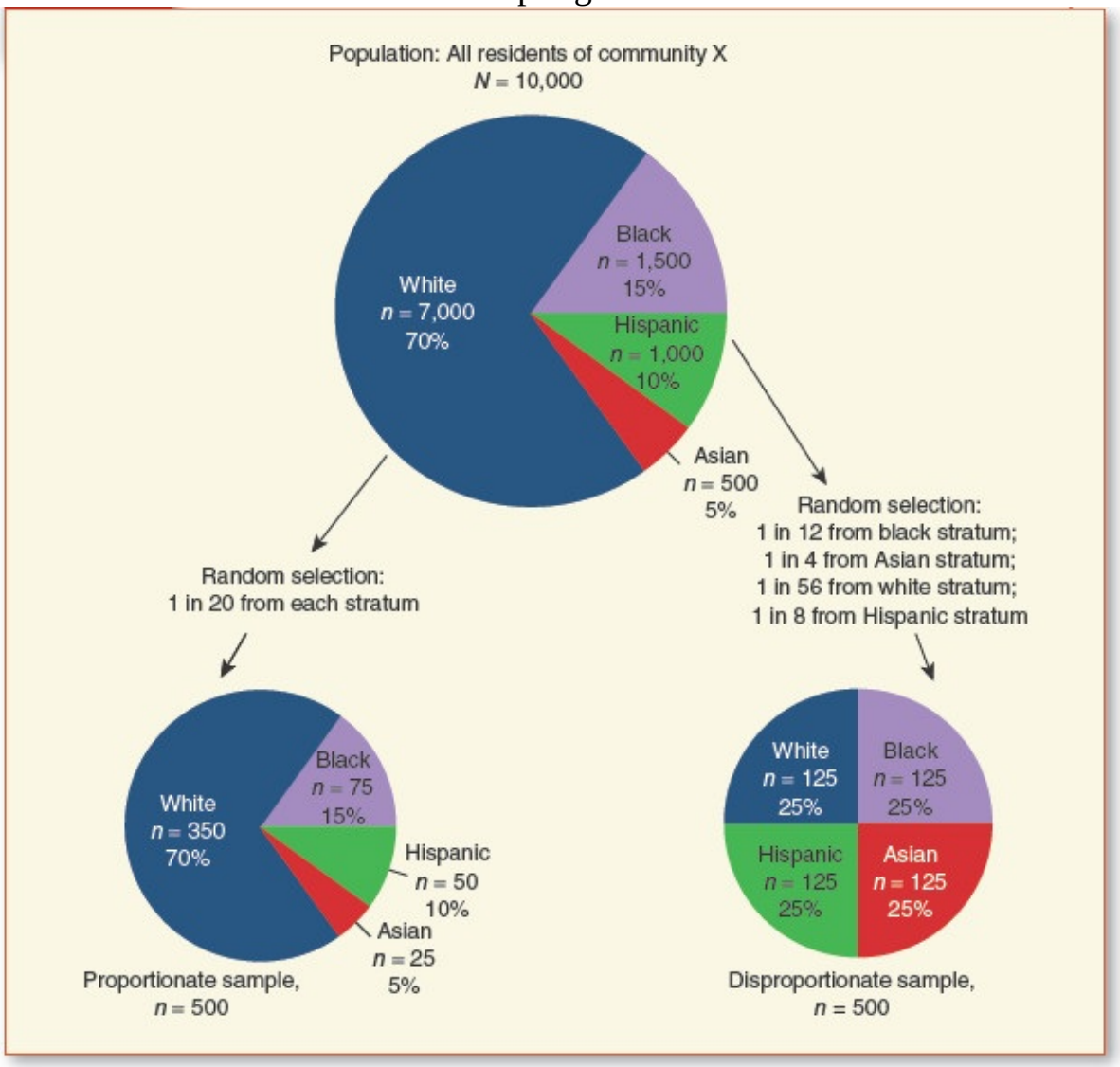
Suppose you want to survey soldiers of an army to determine their morale. Simple random sampling would produce large numbers of enlisted personnel—that is, of lower ranks—but very few, if any, generals. But you want generals in your sample. **Stratified random sampling** ensures that various groups will be included.

First, all elements in the population (that is, in the sampling frame) are distinguished according to their value on some relevant characteristic (army rank, for instance: generals, captains, privates, etc.). That characteristic determines the sampling strata. Next, elements are sampled randomly from within these strata: so many generals, so many captains, and so on. Of course, to use this method, more information is required before sampling than is the case with simple random sampling. Each element must belong to one and only one stratum.

For *proportionate to size* sampling, the size of each stratum in the population must be known. This method efficiently draws an appropriate representation of elements across strata. Imagine that you plan to draw a sample of 500 from an ethnically diverse neighborhood. The neighborhood population is 15% black, 10% Hispanic, 5% Asian, and 70% white. If you drew a simple random sample,

you might end up with somewhat disproportionate numbers of each group. But if you created sampling strata based on race and ethnicity, you could randomly select cases from each stratum in exactly the same proportions. This is termed **proportionate stratified sampling**, and it eliminates any possibility of sampling error in the sample's distribution of ethnicity. Each stratum would be represented exactly in proportion to its size in the population from which the sample was drawn (Exhibit 5.6).

Exhibit 5.6 Stratified Random Sampling



Source: Gallup Organization. 2011. Election polls—Accuracy record in presidential elections. http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print (accessed

March 17, 2011)

In **disproportionate stratified sampling**, the proportion of each stratum that is included in the sample is intentionally varied from what it is in the population. In the case of the sample stratified by ethnicity, you might select equal numbers of cases from each racial or ethnic group: 125 blacks (25% of the sample), 125 Hispanics (25%), 125 Asians (25%), and 125 whites (25%). In this type of sample, the probability of selection of every case is known but unequal between strata. You know what the proportions are in the population, so you can easily adjust your combined sample statistics to reflect these true proportions. For instance, if you want to combine the ethnic groups and estimate the average income of the total population, you would have to weight each case in the sample to reflect its representation in the population.

Why would anyone select a sample that is so unrepresentative in the first place? The most common reason is to ensure that cases from smaller strata are included in the sample in sufficient numbers to allow separate statistical estimates and to facilitate comparisons between strata. Remember that one of the determinants of sample quality is sample size. The same is true for subgroups within samples. If a key concern in a research project is to describe and compare the incomes of people from different racial and ethnic groups, then it is important that the researchers base the mean income of each group on enough cases to be a valid representation. If few members of a particular minority group are in the population, they need to be oversampled.

Stratified random sampling: A method of sampling in which sample elements are selected separately from population strata that the researcher identifies in advance.

Proportionate stratified sampling: Sampling method in which elements are selected from strata in exact proportion to their representation in the population.

Disproportionate stratified sampling: Sampling in which elements are selected from strata in proportions different from those that appear in the population.

Nonprobability Sampling Methods

Nonprobability sampling methods are often used in qualitative research; they also are used in quantitative studies when researchers are unable to use probability selection methods. There are four common nonprobability sampling methods: (1) availability sampling, (2) quota sampling, (3) purposive sampling, and (4) snowball sampling. Because they do not use a random selection procedure, we cannot expect a sample selected with any of these methods to yield a representative sample. Nonetheless, these methods are useful when random sampling is not possible; with a research question that calls for an intensive investigation of a small population; or for a preliminary, exploratory study.

Encyclopedia Link

Read an overview of nonprobability sampling methods.

Availability Sampling

Elements are selected for **availability sampling** (sometimes called *convenience sampling*) because they're available or easy to find. For example, sometimes people stand outside stores in a shopping mall asking passersby to answer a few questions about their shopping habits. That may make sense, but asking the same people for their views on the economy doesn't. In certain respects, regular mall shoppers are not representative people.

An availability sample is often appropriate at key points in social research—for example, when a field researcher explores a new setting and tries to get some sense of prevailing attitudes or when a survey researcher conducts a preliminary test of a new set of questions. If representativeness is not really your goal, availability sampling could be fine. It may be adequate, for instance, when your purpose is really to just make respondents feel appreciated—customers in a store, say, or if you're doing a class project where you're just learning to use a survey or do interviews. Intensive qualitative research efforts, focused less on generalizability than on internal validity, also often rely on availability samples.

Howard Becker’s (1963) classic work on jazz musicians, for instance, was based on groups in which Becker himself played.

Audio Link

Listen to an account of a study using a convenience sampling method.

Availability sampling often masquerades as a more rigorous form of research. Popular magazines periodically survey their readers by printing a questionnaire for readers to fill out and mail in. For many years, *Playboy* magazine has conducted a sex survey among its readers using this technique. But usually only a small fraction of readers return the questionnaire, and these respondents might —how to say it?—have more interesting sex lives than other readers of *Playboy* have, not to mention the rest of us (or so they claim).

Interactive Exercises

Identifying Sampling Techniques

Availability sampling: Sampling in which elements are selected on the basis of convenience.

Careers and Research

Ross Koppel, PhD, Sociologist

Source: Ross Koppel

Sociologist Ross Koppel received his BA, MA, and PhD at Temple University in Philadelphia. In 1985, he founded the Social Research Corporation and since then has served as SRC's president. His work has had major impacts across society. One of his most ambitious research projects was developed initially in response to a request to study the Boston public transit system's (MBTA) treatment of people with disabilities. In 2010, he received the American Sociological Association Distinguished Career Award for the Practice of Sociology.

Koppel's (2008: 11–13) Boston public transit system study involved a unique sampling design. A spreadsheet of all scheduled bus routes allowed him to randomly sample routes and locations. Persons with disabilities who navigated with wheelchairs, walkers, or canes were trained as research observers and then sent to selected routes. The observers rode the selected bus routes and recorded in total almost 1,000 hours of observations of people in wheelchairs and with walkers or canes using buses and the problems they encountered.

Quota Sampling

Quota sampling is intended to overcome the most obvious flaw of availability

sampling—that the sample will just consist of whoever or whatever is available, whether or not it represents the population. In this approach, quotas are set to ensure that the sample represents certain characteristics in proportion to their prevalence in the population, especially if you already know that those characteristics are crucial.

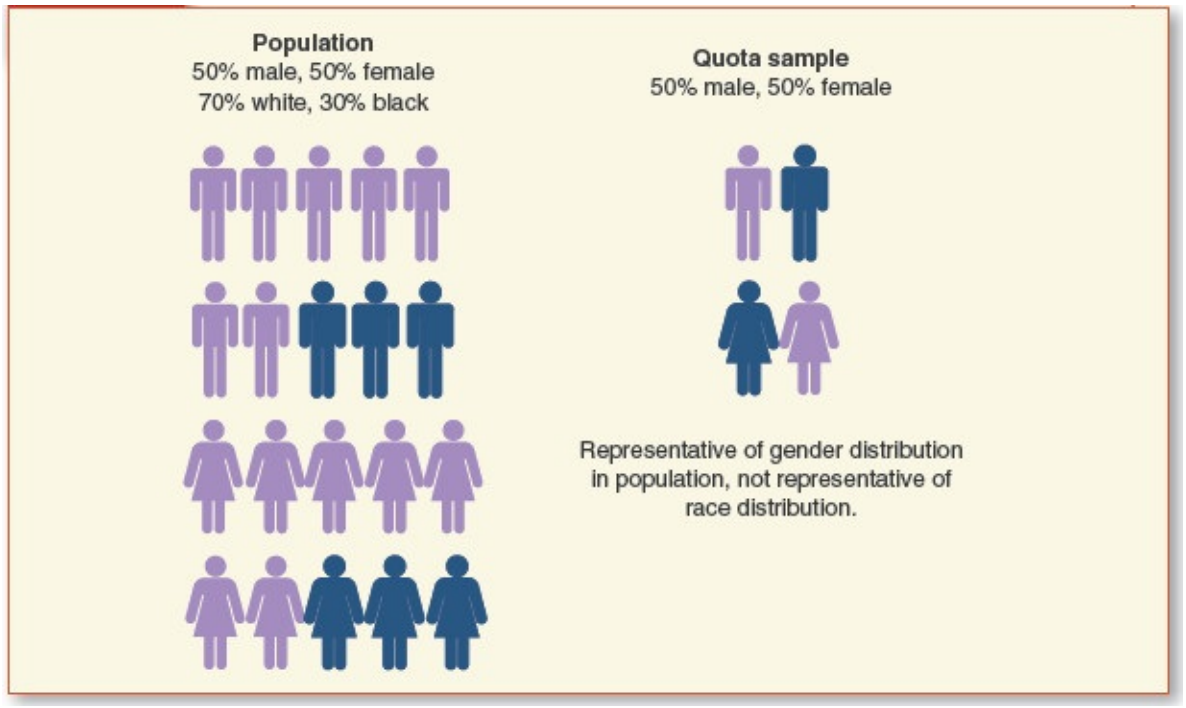
Suppose that you want to sample 500 adult residents of a town. You know from the town's annual report what the proportions of town residents are in gender, employment status, and age. To draw a quota sample of a certain size, you then specify that interviews must be conducted with 500 residents who match the town population in terms of gender, employment status, and age.

The problem is that even when we know that a quota sample is representative of the particular characteristics for which quotas have been set, we have no way of knowing if the sample is representative for any other characteristics. In Exhibit 5.7, for example, quotas have been set for gender only. Under the circumstances, it's no surprise that the sample is representative of the population only for gender, not for race.

Of course, you must know the relevant characteristics of the entire population to set the right quotas. In most cases, researchers know what the population looks like in terms of no more than a few of the characteristics relevant to their concerns. And in some cases, they have no such information on the entire population.

If you're now feeling skeptical of quota sampling, you've gotten the drift of our remarks. Nonetheless, in situations in which you can't draw a random sample, it may be better to establish quotas than to have no parameters at all.

Exhibit 5.7 Quota Sampling


Population
50% male, 50% female
70% white, 30% black
Quota sample
50% male, 50% female
Representative of gender distribution in population, not representative of race distribution.

Quota sampling: A nonprobability sampling method in which elements are selected to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Purposive Sampling

In **purposive sampling**, each sample element is selected for a purpose, usually because of the unique position of the sample elements. Purposive sampling may involve studying the entire population of some limited group (directors of shelters for homeless adults) or a subset of a population (mid-level managers with a reputation for efficiency). Or a purposive sample may be a *key informant survey*, which targets individuals who are particularly knowledgeable about the issues under investigation.

Herbert Rubin and Irene Rubin (1995) suggest three guidelines for selecting informants when designing any purposive sampling strategy. Informants should be

1. Knowledgeable about the cultural arena, situation, or experience being studied
2. Willing to talk
3. Representative of the range of points of view (p. 66)

In addition, Rubin and Rubin (1995) suggest continuing to select interviewees until you can pass two tests:

1. Completeness—"What you hear provides an overall sense of the meaning of a concept, theme, or process."
2. Saturation—"You gain confidence that you are learning little that is new from subsequent interview[s]." (pp. 72–73)

Research|Social Impact Link

Read about one controversy the US Census Bureau deals with in sampling.

Adhering to these guidelines will help ensure that a purposive sample adequately represents the setting or issues studied.

Of course, purposive sampling does not produce a sample that represents some larger population, but it can be exactly what is needed in a case study of an organization, community, or some other clearly defined and relatively limited group.

> **Purposive sampling**: A nonprobability sampling method in which elements are selected for a purpose, usually because of their unique position.

Snowball Sampling

For **snowball sampling**, you identify and speak to one member of the population, and then ask that person to identify others in the population and speak to them, then ask them to identify others, and so on. The sample thus "snowballs" in size. This technique is useful for hard-to-reach or hard-to-identify, interconnected populations (at least some members of the population know each other). An example of a study using snowball sampling is Patricia Adler's (1993) study of Southern California drug dealers. Wealthy philanthropists, top business executives, or Olympic athletes, all of who may have reason to refuse a "cold call" from an unknown researcher, might be sampled effectively using the snowball technique. However, researchers using snowball sampling normally cannot be confident that their sample represents the total population of interest, so generalizations must be tentative.

Snowball sampling: A method of sampling in which sample elements are selected as successive informants or interviewees identify them.

Conclusion

Sampling is a powerful tool for social science research. Probability sampling methods allow a researcher to use the laws of chance, or probability, to draw samples from which population parameters can be estimated with a high degree of confidence. A sample of just 1,000 or 1,500 individuals can be used to estimate reliably the characteristics of the population of a nation comprising millions of individuals.

But researchers do not come by representative samples easily. Well-designed samples require careful planning, some advance knowledge about the population to be sampled, and adherence to systematic selection procedures—all so that the selection procedures are not biased. And even after the sample data are collected, the researcher's ability to generalize from the sample findings to the population is not completely certain.

The alternatives to random, or probability-based, sampling methods are almost always much less palatable for quantitative studies, even though they are typically much cheaper. Without a method of selecting cases likely to represent the population in which the researcher is interested, research findings must be carefully qualified. Qualitative researchers whose goal is to understand a small group or setting in depth may necessarily have to use unrepresentative samples, but they must keep in mind that the generalizability of their findings will not be known. Additional procedures for sampling in qualitative studies will be introduced in Chapter 9.

Social scientists often seek to generalize their conclusions from the population that they studied to some larger target population. Careful design of appropriate sampling strategies makes such generalizations possible.

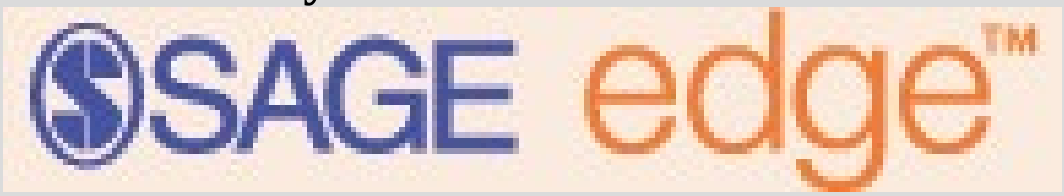
Key Terms

Availability sampling 103
Bias 96
Census 94
Cluster 100
Cluster sampling 99
Disproportionate stratified sampling 102
Elements 92
Nonprobability sampling method 96
Periodicity 99
Population 91
Probability of selection 96
Probability sampling method 96
Proportionate stratified sampling 101
Purposive sampling 105
Quota sampling 104
Random digit dialing (RDD) 99
Random number table 99
Random sampling 96
Representative sample 94
Sample 92
Sampling frame 92
Sampling interval 99
Sampling units 92
Simple random sampling 99
Snowball sampling 105
Stratified random sampling 101
Systematic random sampling 99
Target population 94

Highlights

- Sampling theory focuses on the generalizability of descriptive findings to the population from which the sample was drawn. It also considers whether statements can be generalized from one population to another.
- Sampling is unnecessary when the elements that would be sampled are identical, but the complexity of the social world often makes it difficult to argue that different elements are identical. Conducting a complete census of a population also eliminates the need for sampling, but the resources required for a complete census of a large population are usually prohibitive.
- Nonresponse undermines sample quality: The obtained sample, not the desired sample, determines sample quality.
- Probability sampling methods rely on a random selection procedure to ensure no systematic bias in the selection of elements. In a probability sample, the odds of selecting elements are known, and the method of selection is carefully controlled. This should result in a representative sample, in which the selection of elements is unbiased.
- A sampling frame (a list of elements in the population) is required in most probability sampling methods. The adequacy of the sampling frame is an important determinant of sample quality.
- Simple random sampling and systematic random sampling are equivalent probability sampling methods in most situations. However, systematic random sampling is inappropriate for sampling from lists of elements that have a regular, periodic structure.
- Cluster sampling is less efficient than simple random sampling but is useful when a sampling frame is unavailable. It is also useful for large populations spread out across a wide area or among many organizations.
- Stratified random sampling uses prior information about a population to make sampling more efficient. Stratified sampling may be either proportionate or disproportionate. Disproportionate stratified sampling is useful when a research question focuses on a stratum or on strata that make up a small proportion of the population.
- Nonprobability sampling methods can be useful when random sampling is not possible, when a research question does not concern a larger population, and when a preliminary exploratory study is appropriate. However, the representativeness of nonprobability samples cannot be determined.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. When (if ever) is it reasonable to assume that a sample is not needed because “everyone is the same”—that is, the population is homogeneous? Does this apply to research such as Stanley Milgram’s on obedience to authority? What about investigations of student substance abuse? How about investigations of how people (or their bodies) react to alcohol? What about research on likelihood of voting (the focus of Chapter 8)?
2. All adult U.S. citizens are required to participate in the decennial census, but some do not. Some social scientists have argued for putting more resources into getting a large representative sample so that census takers can secure higher rates of response from hard-to-include groups. Do you think that the U.S. Census should shift to a probability-based sampling design? Why or why not?
3. What increases sampling error in probability-based sampling designs? Stratified rather than simple random sampling? Disproportionate (rather than proportionate) stratified random sampling? Stratified rather than cluster random sampling? Why do researchers select disproportionate (rather than proportionate) stratified samples? Why do they select cluster rather than simple random samples?
4. What are the advantages and disadvantages of probability-based sampling designs compared with nonprobability-based designs? Could any of the researches that are described in this chapter with a nonprobability-based design have been conducted instead with a probability-based design? What difficulties might have been encountered in an attempt to use random selection? How would you discuss the degree of confidence you can place in the results obtained from research using a nonprobability-based sampling design?

Finding Research

1. Locate one or more newspaper articles reporting the results of an opinion poll. What information does the article provide on the sample that was selected? What additional information do you need to determine whether the sample was a representative one?
2. From professional journals, select five articles that describe research using a sample drawn from some population. Identify the type of sample used in each study, and note any strong and weak points in how the sample was actually drawn. Did the researchers have a problem resulting from nonresponse? Considering the sample, how confident are you in the validity of generalizations about the population based on the sample? Do you need any additional information to evaluate the sample? Do you think a different sampling strategy would have been preferable? To what larger population were the findings generalized? Do you think these generalizations were warranted? Why or why not?
3. Research on time use has been flourishing all over the world in recent years. Search the web for sites that include the words *time use* and see what you find. Choose one site and write a paragraph about what you learned from it.
4. Check out the “People and Households” section of the U.S. Census Bureau website (www.census.gov). Based on some of the data you find there, write a brief summary of some aspect of the current characteristics of the U.S. population.

Critiquing Research

1. Shere Hite's popular book *Women and Love* (1987) is a good example of the claims that are often made based on an availability sample. In this case, however, the sample didn't necessarily appear to be an availability sample because it consisted of so many people. Hite distributed 100,000 questionnaires to church groups and many other organizations and received back 4.5%; 4,500 women took the time to answer some or all of her 127 essay questions regarding love and sex. Is Hite's sample likely to represent American women in general?

   Why or why not? You might look at the book's empirical generalizations and consider whether they are justified.
2. In newspapers or magazines, find three examples of poor sampling, where someone's conclusions—either in formal research or in everyday reasoning—are weakened by the selection of cases the author has looked at. How is the author's sampling flawed, and how might that systematically distort the findings? Don't just say, "the cases might not be typical"—try to guess, for instance, the direction of error. For example, did the person pick unusually friendly or accessible people? The most well-known examples? And how might their approach affect the findings?

Doing Research

1. Select a random sample using a table of random numbers (either one provided by your instructor or one from a website, such as www.bmra.com/extras/man-rand.htm). Compute a statistic based on your sample and compare it with the corresponding figure for the entire population. Here's how to proceed:
    1. First, select a very small population for which you have a reasonably complete sampling frame. One possibility would be the listing of some characteristic of states in a U.S. Census Bureau publication, such as average income or population size. Another possible population would be the list of asking prices for houses advertised in your local paper.
    2. Next, create a sampling frame, a numbered list of all the available elements in the population. If you are using a complete listing of all elements, as from a U.S. Census Bureau publication, the sampling frame is the same as the list. Just number the elements (states). If your population is composed of housing ads in the local paper, your sampling frame will be those ads that contain a housing price. Identify these ads, and then number them sequentially, starting with 1.
    3. Decide on a method of picking numbers out of the random number table, such as taking every number in each row, row by row, or moving down or diagonally across the columns. Use only the first (or last) digit in each number if you need to select 1 to 9 cases or only the first (or last) two digits if you want 10 to 99 cases.
    4. Pick a starting location in the random number table. It's important to pick a starting point in an unbiased way, perhaps by closing your eyes and then pointing to some part of the page.
    5. Record the numbers you encounter as you move from the starting location in the direction you decided on in advance, until you have recorded as many random numbers as the number of cases you need in the sample. If you are selecting states, 10 might be a good number. Ignore numbers that are too large (or small) for the range of numbers used to identify the elements in the population. Discard duplicate numbers.
    6. Calculate the average value in your sample for some variable that was measured (for example, population size in a sample of states or housing price for the housing ads). Calculate the average by adding the values of all the elements in the sample and dividing by the number of elements in the sample.
    7. Go back to the sampling frame and calculate this same average for all the elements in the list. How close is the sample average to the population average?
    8. Estimate the range of sample averages that would be likely to include 90% of the possible samples.

Ethics Questions

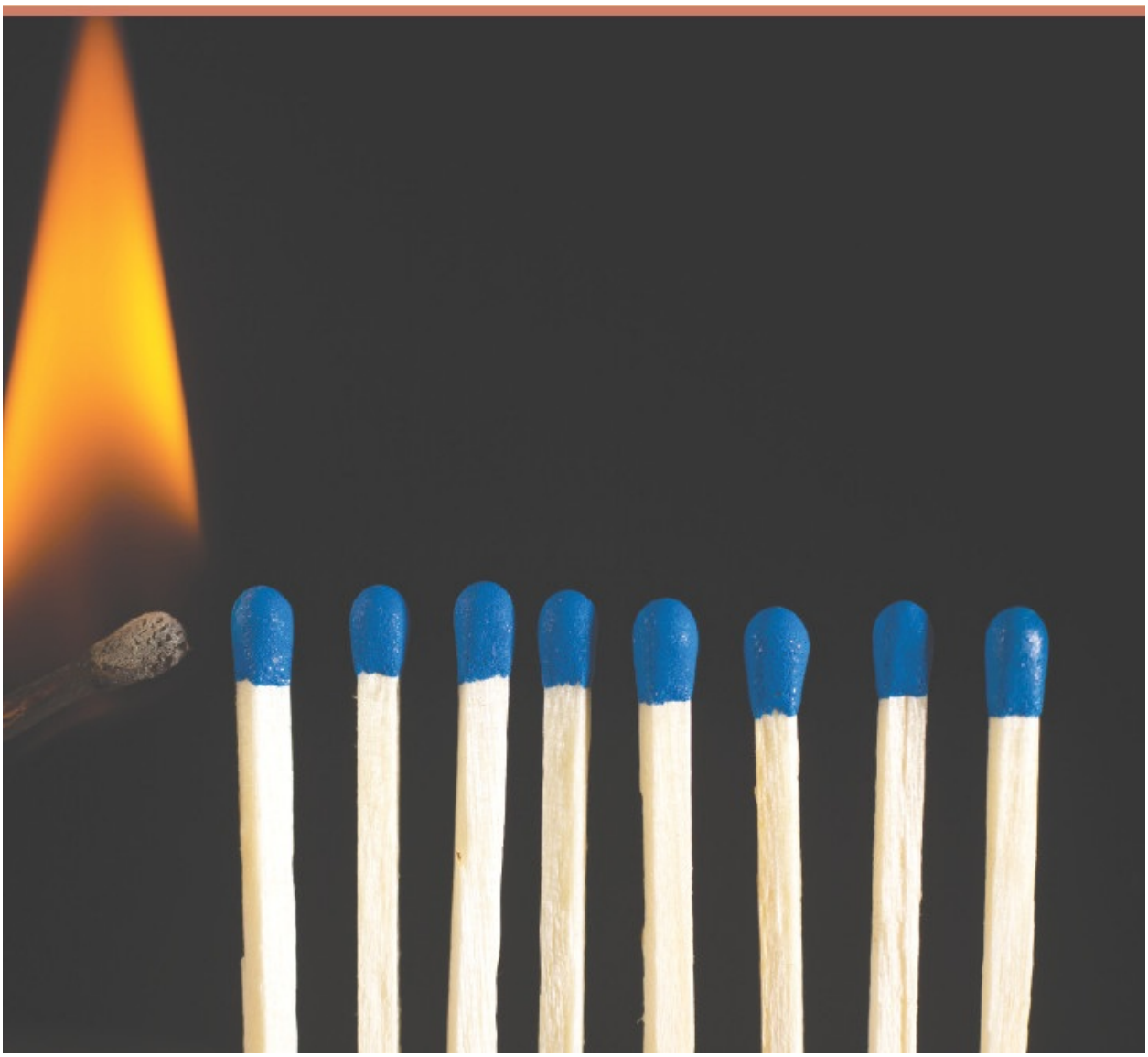
1. How much pressure is too much pressure to participate in a probability-based sample survey? Is it okay for the U.S. government to mandate legally that all citizens participate in the decennial census? Should companies be able to require employees to participate in survey research about work-related issues? Should students be required to participate in surveys about teacher performance? Should parents be required to consent to the participation of their high school–age students in a survey about substance abuse and health issues? Is it okay to give monetary incentives for participation in a survey of homeless shelter clients? Can monetary incentives be coercive? Explain your decisions.
2. Federal regulations require special safeguards for research on persons with impaired cognitive capacity. Special safeguards are also required for research on prisoners and on children. Do you think special safeguards are necessary? Why or why not? Do you think it is possible for individuals in any of these groups to give “voluntary consent” to research participation? What procedures might help make consent to research truly voluntary in these situations? How could these procedures influence sampling plans and results?

Video Interview Questions

Listen to the researcher interview for Chapter 5 at edge.sagepub.com/chamblissmssw5e.

1. What was Anthony Roman's research question in his phone survey research study?
2. What were Roman's major discoveries in this project? How does this emphasize the importance of sampling selectively and carefully?

Chapter 6 Causation and Experimental Design



Burcak Yalaman/iStock/Thinkstock

Learning Objectives

1. List the five criteria for establishing a causal relationship.
2. Explain the meaning of the expression "correlation does not prove causation."
3. Compare the difference between an independent and a dependent variable and understand their function in an experiment
4. List the essential components of a true experimental research design.
5. Distinguish the concepts of random assignment (randomization) and random sampling.
6. Identify the two major types of quasi-experimental design and explain why they are considered to be "quasi" experimental.
7. Define "ex post facto control group design" and explain why it is not considered to be experimental or quasi-experimental.
8. Discuss the influences on external validity (generalizability) in experimental design and those on internal validity (causal validity).
9. Explain the role of process analysis in experimental research.
10. Discuss the most distinctive ethical challenges in experimental research.

Identifying causes—figuring out why things happen—is the goal of most social science research. Unfortunately, valid explanations of the causes of social phenomena do not come easily. Why did the homicide rate in the United States drop for 15 years and then start to rise in 1999 (Butterfield 2000: 12)? Was it because of changes in the style of policing (Radin 1997) or because of changing attitudes among young people (Butterfield 1996a)? Was it the result of variation in patterns of drug use (Krauss 1996), or to more stringent handgun regulations (Butterfield 1996b)? Did better emergency medical procedures result in higher survival rates for victims (Ramirez 2002)? If we are to evaluate these alternative explanations, we must design our research strategies carefully.

This chapter considers the meaning of causation, the criteria for achieving causally valid explanations, the ways in which experimental and quasi-experimental research designs seek to meet these criteria, and the difficulties that can sometimes result in invalid conclusions. By the end of the chapter, you should have a good grasp of the meaning of causation and the logic of experimental design. Most social research, both academic and applied, uses data collection methods other than experiments. But because experimental designs are the best way to evaluate causal hypotheses, a better understanding of them will help you to be aware of the strengths and weaknesses of other research

designs, which we will consider in subsequent chapters.

Causal Explanation

A cause is an explanation of some characteristic, attitude, or behavior of groups, individuals, or other entities (such as families, organizations, or cities) or of events. For example, Lawrence Sherman and Richard Berk (1984) conducted a study to determine whether adults who were accused of a domestic violence offense would be less likely to repeat the offense if police arrested them rather than just warned them. Sherman and Berk's conclusion that this hypothesis was correct meant that they believed police response had a causal effect on the likelihood of committing another domestic violence offense.

Video Link

Watch an overview on carefully planning your social research.

More specifically, a **causal effect** is said to occur if variation in the independent variable is followed by variation in the dependent variable, when all other things are equal—***ceteris paribus***. For instance, we know that for the most part, men earn more income than women do. But is this because they are men—or could it result from higher levels of education or from longer tenure in their jobs (with no pregnancy breaks), or is it because of the kinds of jobs men go into compared with those that women choose? We want to know if men earn more than women, *ceteris paribus*—other things (job, tenure, education, etc.) being equal.

Of course, "all" other things can't literally be equal: We can't compare the same people at the same time in exactly the same circumstances except for the variation in the independent variable (King, Keohane, & Verba 1994). However, we can design research to create conditions that are comparable so that we can isolate the impact of the independent variable on the dependent variable.

Causal effect: The finding that change in one variable leads to change in another variable, *ceteris paribus* (other things being equal). *Example:* Individuals arrested for domestic assault tend to commit fewer subsequent assaults than similar individuals who are accused in the same circumstances but are not arrested.

Ceteris paribus: Latin phrase meaning "other things being equal."

What Causes What?

Five criteria should be considered in trying to establish a causal relationship. The first three criteria are generally considered as requirements for identifying a causal effect: (1) empirical association, (2) temporal priority of the independent variable, and (3) nonspuriousness. You must establish these three to claim a causal relationship. Evidence that meets the other two criteria—(4) identifying a causal mechanism and (5) specifying the context in which the effect occurs—can considerably strengthen causal explanations.

Journal Link

Read an article that looks at a potential cause of delinquency.

Research designs that allow us to establish these criteria require careful planning, implementation, and analysis. Many times, researchers have to leave one or more of the criteria unmet and are left with some important doubts about the validity of their causal conclusions, or they may even avoid making any causal assertions.

Association

The first criterion for establishing a causal effect is an empirical (or observed) **association** (sometimes called a *correlation*) between the independent and dependent variables. They must vary together such that when one goes up (or down), the other goes up (or down) at the same time. Here are some examples: When cigarette smoking goes up, so does lung cancer. The longer you stay in school, the more money you will make later in life. Single women are more likely to live in poverty than are married women. When income goes up, so does overall health. In all of these cases, a change in an independent variable correlates, or is associated with, a change in a dependent variable. If there is no association, there cannot be a causal relationship. For instance, empirically there seems to be no correlation between the use of the death penalty and a reduction in the rate of serious crime. That may seem unlikely to you, but empirically it is the case: There is no correlation. So there cannot be a causal relationship.

Association: A criterion for establishing a causal relationship between two variables: Variation in one variable is empirically related to variation in another variable.

Research in the News

Where Did You Hear That?

Misinformation absorbed from media can have lasting impacts on perspective. Professor of communications Jakob Jensen used a media experiment to test the persistence of media messages in our minds. Jensen used a quasi-experimental design in which students watched a TV episode. He then tested their beliefs related to events in the TV episode right after watching the show and then two weeks later. He found that the show's message was actually more pronounced among participants two weeks after watching the show than immediately after it.

1. Identify a TV show that you think might have lasting effects on student orientations to an issue that concerns you. Explain your reasoning.
2. What causal mechanism might explain the greater long-term impact of the effect that Jensen identified compared with its short-term impact? How would you design a study to investigate this mechanism?

News source: Cook, Gareth. 2011. TV's sleeper effect. *Boston Globe*, October 30.

Time Order

Association is necessary for establishing a causal effect, but it is not sufficient. We must also ensure that the change in the independent variable came before change in the dependent variable—the cause must come before its presumed effect. This is the criterion of **time order**, or the temporal priority of the independent variable. Motivational speakers sometimes say that to achieve success (the dependent variable in our terms), you really need to believe in yourself (the independent variable). And it is true that many very successful politicians, actors, and businesspeople seem remarkably confident—there is an association. But it may well be that their confidence is the result of their success, not its cause. Until you know which came first, you can't establish a causal connection.

Time order: A criterion for establishing a causal relationship between two variables: The variation in the presumed cause (the independent variable) must occur before the variation in the presumed effect (the dependent variable).

Nonspuriousness: A criterion for establishing a causal relation between two variables; when a relationship between two variables is not caused by variation in a third variable.

Spurious: Nature of a presumed relationship between two variables that actually results from variation in a third variable.

Nonspuriousness

The third criterion for establishing a causal effect is **nonspuriousness**. *Spurious* means false or not genuine. We say that a relationship between two variables is **spurious** when it is actually caused by changes in a third variable, so what appears to be a direct connection is in fact not one. Have you heard the old adage "Correlation does not prove causation"? It is meant to remind us that an association between two variables might be caused by something else. If we measure children's shoe sizes and their academic knowledge, for example, we will find a positive association. However, the association results from the fact that older children have larger feet as well as more academic knowledge. A third variable (age) is affecting both shoe size and knowledge so that they correlate, but one doesn't cause the other. Shoe size does not cause knowledge, or vice versa. The association between the two is, we say, spurious.

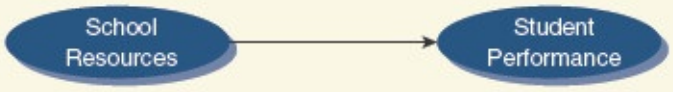
If this point seems obvious, consider a social science example. Do schools with better resources produce better students? There is certainly a correlation, but consider the fact that parents with more education and higher income tend to live in neighborhoods that spend more on their schools. These parents are also more likely to have books in the home and to provide other advantages for their children (Exhibit 6.1). Maybe parents' income causes variation in both school resources and student performance. If so, there would be an association between school resources and student performance, but it would be at least partially spurious. What we want, then, is nonspuriousness.

Encyclopedia Link

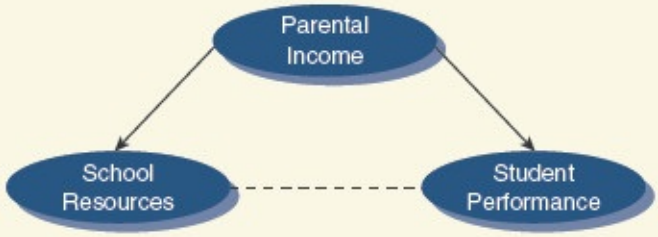
Read more about spurious correlations.

Exhibit 6.1 A Spurious Relationship Revealed

School resources are associated with student performance; apparently, a causal relation.



But in fact, parental income (a third variable) influences both school resources and student performance, creating the association.



Research That Matters

A popular theory says that economic distress causes crime. But since 2005, although youth unemployment in the United Kingdom has been increasing, "youth offending [has been] in sharp and sustained decline" at the same time (Fergusson 2013: 31). Ross Fergusson (2103: 52) at the UK's Open University was puzzled by this pattern and decided to conduct an extensive review of prior research to better understand these "potentially contradictory issues" about the causes of youth crime.

Fergusson found that research conclusions about the unemployment–crime association were complex, varying with the type of crime measured, the ages of youth studied, and the use of aggregate or individual data. He also remained unconvinced that new crime-prevention programs had been responsible for the unexpected decline in crime. He concluded that the criminogenic effects of unemployment could be delayed or that they could be displaced by a turn toward mass protests.

Source: Adapted from Fergusson, Ross. 2013. Risk, responsibilities and rights: Reassessing the "economic causes of crime" thesis in a recession. *Youth Justice* 13(1): 31–56.

Mechanism

A causal **mechanism** is the process that creates the connection between the variation in an independent variable and the variation in the dependent variable that it is hypothesized to cause (Cook & Campbell 1979: 35; Marini & Singer 1988). Many social scientists (and scientists in other fields) argue that no causal explanation is adequate until a mechanism is identified.

For instance, there seems to be an empirical association at the individual level between poverty and delinquency: Children who live in impoverished homes seem more likely to be involved in petty crime. But why? What is the mechanism by which living in these homes "produces" petty crime? Some researchers have argued for a *mechanism* of low parent–child attachment, inadequate supervision of children, and erratic discipline as the means by which poverty and delinquency are connected (Sampson & Laub 1994). Or a different example: It's clearly true that religion affects adolescent sexual attitudes and behavior. But how does this work? The answer seems to lie in some combination of religious morality (beliefs), involvement (for instance, spending time in church activities keeps teenagers from having the time for sexual adventures), and religious subcultures (for instance, having peer relationships that discourage sex). In trying to distinguish the impact of these, researchers are looking for the mechanism by which "religion" (in some sense!) affects sexual behavior (Regnerus 2007).

Figuring out how the process by which the independent variable influenced the variation in the dependent variable can increase confidence in our conclusion that a causal effect was at work (Costner 1989).

Mechanism: A discernible process that creates a causal connection between two variables.

Context

No cause has its effect apart from some larger **context**. When, for whom, and in what conditions does this effect occur? A cause is really one among a set of interrelated factors required for the effect (Hage & Meeker 1988; Papineau 1978). Identification of the context is not itself a criterion for a valid causal conclusion, but it does help us to understand the relationship and when it applies.

You may hypothesize, for example, that if you offer employees higher wages to work harder, they will indeed work harder. In the context of capitalist America, this seems indeed to be the case; incentive pay causes harder work. But in noncapitalist societies, workers often want only enough money to meet their basic needs and would rather work less than drive themselves hard just to have more money. In the United States, the correlation of incentive pay with greater effort seems to work; in medieval Europe, for instance, it did not (Weber 1930/1992).

Research|Social Impact Link

Read a critique about a potential spurious relationship within scientific research.

Or to return to the juvenile justice example, Robert Sampson and John Laub (1993) looked at 538,000 cases ranging across 322 U.S. counties, and found that context—that is, where the cases happened—mattered quite a lot. In counties with a relatively large underclass and a concentration of poverty among minorities, juvenile offenders were treated more harshly than in more prosperous areas. This effect occurred among both whites and African Americans, but it was particularly strong among the African Americans.

A particular historical period can also be an important context for research findings. In the United States during the 1960s, for instance, children of divorced parents (“from a broken home, ” as the expression was then) were more likely to suffer from a variety of problems; crucially, they lived in a context of mostly intact families. In recent years, though, many parents are divorced, and the causal link between divorced parents and social pathology no longer seems to hold (Coontz 1997).

Context: The larger set of interrelated circumstances in which a particular outcome should be understood.

Why Experiment?

You can see, then, that establishing a causal relationship between two variables can be quite difficult. The “gold standard” for scientific research design, as you may know, is conducting experiments. Experiments provide the most powerful design for testing causal hypotheses because they allow us to establish confidently the first three criteria for causality—association, time order, and nonspuriousness. So-called **true experiments** have at least three features that help us meet these criteria:

1. Two comparison groups (in the simplest case, an experimental group and a control group), which establish association
2. Variation in the independent variable before assessment of change in the dependent variable, which establishes time order
3. Random assignment to the two (or more) comparison groups, which establishes nonspuriousness

We can determine whether an association exists between the independent and dependent variables in a true experiment because two or more groups, the **comparison groups**, differ in their value on the independent variable. One group receives some “treatment,” which is a manipulation of the value of the independent variable. This group is termed the **experimental group**. In a simple experiment, there may be one other group that does not receive the treatment; it is called a **control group**.

Consider an example in detail (Exhibit 6.2). Does drinking coffee improve one’s writing of an essay? Imagine a simple experiment. Suppose you believe that drinking two cups of strong coffee before class will help you in writing an in-class essay. But other people think that coffee makes them too nervous and “wired” and so doesn’t help in writing the essay. To test your hypothesis (“Coffee drinking causes improved performance.”), you need to compare two groups of subjects, a control group and an experimental group. First, the two groups will sit and write an in-class essay. Then, the control group will drink no coffee, and the experimental group will drink two cups of strong coffee. Next, both groups will sit and write another in-class essay. At the end, all of the essays will be graded, and you will see which group improved more. Thus, you may establish *association.*

Video Link

Watch a clip about using control groups in current social research.

Exhibit 6.2 A True Experiment

Experimental Group:	R	O_1	X	O_2
Comparison Group:	R	O_1		O_2

Key: R = Random assignment
O = Observation (pretest [O_1] or posttest [O_2])
X = Experimental treatment

	O_1	X	O_2
Experimental Group	Pretest Essay	Coffee	Posttest Essay
Comparison Group	Pretest Essay		Posttest Essay

You may find an association outside such an experimental setting, of course, but it wouldn't establish time order. Perhaps good writers hang out in cafés and only then start drinking lots of coffee. So there would be an association, but not the causal relation we're looking for. By controlling who gets the coffee, and when, we establish *time order.*

All experiments have a **posttest**—that is, a measurement of the outcome in both groups after the experimental group has received the treatment. In our example, you grade the papers. Many true experiments also have **pretests**, which measure the dependent variable before the experimental intervention. A pretest is exactly the same as a posttest, just administered at a different time. Strictly speaking, though, a true experiment does not require a pretest. When researchers use random assignment, the groups' initial scores on the dependent variable and on all other variables are very likely to be similar. Any difference in outcome between the experimental and comparison groups is therefore likely to result from the intervention (or to other processes occurring during the experiment), and the likelihood of a difference just on the basis of chance can be calculated.

Finally, remember that the two groups must be as equal as possible at the beginning of the study. If you let students choose which group to be in, ambitious students may pick the coffee group, hoping to stay awake and do better on the paper. Or, people who simply don't like the taste of coffee may choose the noncoffee group. Either way, your two groups won't be equivalent at the beginning of the study, and any difference in their writing may be the result of that initial difference (a source of spuriousness), not the drinking of coffee. Finally, as our colleague Stan Lieberson has pointed out to us, coffee affects coffee drinkers and nondrinkers quite differently. Ideally, we'd have similar proportions of each in our different groups.

Journal Link

Read an article that uses matching as a way of comparing group treatment.

So, you randomly sort the students into the two different groups. You can do this by flipping a coin for each student, by pulling names out of a hat, or by using a random number table as described in the previous chapter. In any case, the subjects themselves should not be free to choose, nor should you (the experimenter) be free to put them into whatever group you want. (If you did that, you might unconsciously put the better students into the coffee group, hoping to get the results you're looking for.) Thus, we hope to achieve nonspuriousness.

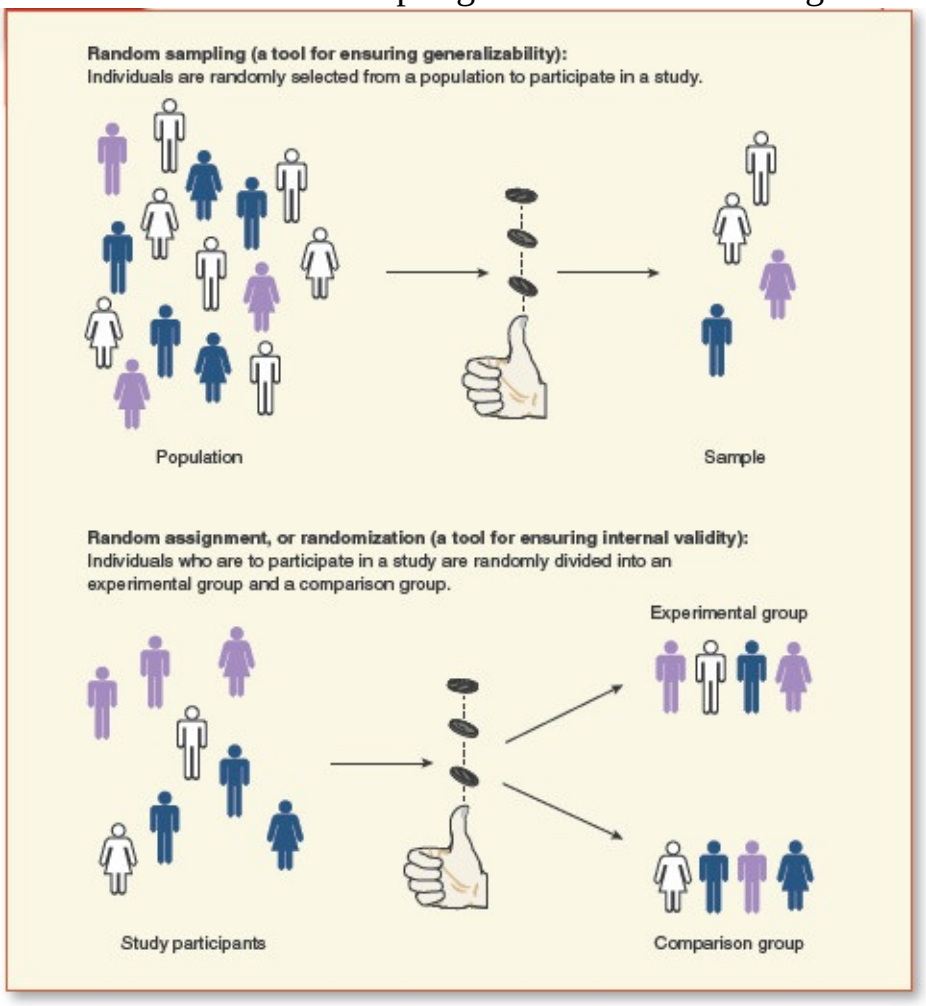
Note that the random assignment of subjects to experimental and comparison groups is not the same as random sampling of individuals from some larger population (Exhibit 6.3). In fact, **random assignment (randomization)** does not help at all to ensure that the research subjects are representative of some larger population; instead, representativeness is the goal of random sampling. What random assignment does—create two (or more) equivalent groups—is useful for ensuring internal validity, not generalizability.

Matching is another procedure sometimes used to equate experimental and comparison groups, but by itself, it is a poor substitute for randomization. Matching of individuals in a treatment group with those in a comparison group might involve pairing persons on the basis of similarity of gender, age, year in school, or some other characteristic. The basic problem is that, as a practical matter, individuals can be matched on only a few characteristics; unmatched differences between the experimental and comparison groups may still influence

outcomes.

These defining features of true experimental designs give us a great deal of confidence that we can meet the three basic criteria for identifying causes: association, time order, and nonspuriousness. However, we can strengthen our understanding of causal connections, and increase the likelihood of drawing causally valid conclusions, by also investigating causal mechanism and causal context.

Exhibit 6.3 Random Sampling Versus Random Assignment


Random sampling (a tool for ensuring generalizability):
Individuals are randomly selected from a population to participate in a study.
Population
Sample
Random assignment, or randomization (a tool for ensuring internal validity):
Individuals who are to participate in a study are randomly divided into an experimental group and a comparison group.
Experimental group
Study participants
Comparison group

When true experiments can be done—a rarity in the "real world," for social science—the resulting knowledge can be quite valuable. In 2008, the state of Oregon was preparing to expand its Medicaid program for low-income families, but only had enough money to cover 10,000 people of the 90,000 who applied. (Finkelstein et al. 2011) The state, aided by a team of social scientists, decided to

run an experiment to see whether Medicaid truly did benefit its recipients. A lottery of the applicants was conducted, with the 10,000 recipients therefore being randomly selected. Within a year, some results were clear: a tremendous reduction in financial hardship, a dramatic reduction in depression, and a 25% improvement in recipients' self-reports of good to excellent health. There was also a clear increase in their use of medical services and facilities, although the results of objective physical health were much more ambiguous. Such studies are very unusual—most of the time, people will not consent to being randomly selected to receive what they believe to be valuable services—but the Oregon Health Insurance Experiment was one of the most scientifically impressive and practically useful studies in many years.

Researcher Interview Link

Watch a researcher elaborate on experimental designs.

True experiment: Experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation. Outcomes are measured in a posttest.

Comparison groups: In an experiment, groups that have been exposed to different treatments or values of the independent variable (e.g., a control group and an experimental group).

Experimental group: In an experiment, the group of subjects that receives the treatment or experimental manipulation.

Control group: A comparison group that receives no treatment.

Posttest: In experimental research, the measurement of an outcome (dependent) variable after an experimental intervention or after a presumed independent variable has changed for some other reason. The posttest is exactly the same "test" as the pretest, but it is administered at a different time.

Pretest: In experimental research, the measurement of an outcome (dependent) variable before an experimental intervention or change in a presumed independent variable for some other reason. The pretest is exactly the same "test" as the posttest, but it is administered at a different time.

Random assignment (randomization): A procedure by which each experimental subject is placed in a group randomly.

Matching: A procedure for equating the characteristics of individuals in different comparison

groups in an experiment. Matching can be done on either an individual or an aggregate basis. For individual matching, individuals who are similar in key characteristics are paired before assignment, and then the two members of each pair are assigned to the two groups. For aggregate matching, groups chosen for comparison are similar in the distribution of key characteristics.

Careers and Research

Sruthi Chandrasekaran, Senior Research Associate

Source: Sruthi Chandrasekaran

Sruthi Chandrasekaran is a senior research associate at J-PAL—The Abdul Latif Jameel Poverty Action Lab that was established at the Massachusetts Institute of Technology but has become a global network of researchers who seek to reduce poverty by ensuring that policy is informed by scientific evidence. J-PAL emphasizes the use of randomized controlled trials to evaluate the impact of social policies.

Chandrasekaran has completed a 5-year integrated master's in economics at the Indian Institute of Technology (IIT) Madras and an MSc in comparative social policy at the University of Oxford. Her most recent project tests the value of performance-based incentives on improving

tuberculosis (TB) reduction efforts by health workers in North Indian slums.

Chandrasekaran's academic training in economics and social policy provided strong qualitative and quantitative research tools, but her interest in having an impact on societal development led to her career. As a field-based researcher, she meets with communities, listens to their perspectives, and proposes interventions. She then takes the lead in ensuring that the intervention follows the study design to the dot, the data collection tools elicit quality responses in an unbiased manner, the survey data is of the highest quality, the cleaning of the data is coherent and methodical, and the analysis is rigorous. Because study results are published in leading academic journals and the policy lessons are disseminated to key stakeholders, it is crucial that the research is well designed and the quality of the data is impeccable.

Chandrasekaran's research training helps her examine issues in an objective manner, develop a logical framework to investigate issues in detail, and understand the story behind the data. She also strives to affect policy design and implementation by sharing what she has learned in the field. Working with data collected about real problems helps make these tasks interesting, exciting, and rewarding.

Chandrasekaran offers some heartfelt advice for students interested in a career involving doing research or using research results:

> Researchers need the ability to study an aspect of a social problem in great detail as well as the flexibility to step back and look at the bigger picture. Consciously training to don both hats is very helpful. The ability to understand field realities is crucial to designing a research question that is grounded as well as one that is useful for policy analysis. Research can at times be painstakingly slow and frustrating, so patience and single-minded focus on the end goal can help one through the tough times. Being aware of competing methodologies and research studies in relevant fields can also be quite useful in understanding the advantages and pitfalls in your own research. If you are inspired to take up research, make sure you choose a field close to your heart since this will be personally and professionally rewarding. If you are unsure, take up an internship or a short-term project to see how much you may enjoy it.

What If a True Experiment Isn’t Possible?

Unfortunately, in social science true experiments are often not feasible, although social psychologists and some market researchers use them often. In many fields, though, true experiment may be too costly or take too long to carry out; it may not be ethical to randomly assign subjects to the different conditions; or the “treatment” events may already have occurred, so it may be too late to conduct a true experiment. Researchers may then instead use **quasi-experimental designs** that retain several components of experimental design but differ in important details.

Audio Link

Listen to this clip about the different elements of experimental design.

In quasi-experimental design, a comparison group is predetermined to be comparable to the treatment group in critical ways, such as being eligible for the same services or being in the same school cohort (Rossi & Freeman 1989: 313). Such research designs are only quasi-experimental, because subjects are not randomly assigned to the comparison and experimental groups. As a result, we cannot be as confident in the comparability of the groups as in true experimental designs. Nonetheless, to term a research design quasi-experimental, we have to be sure that the comparison groups meet specific criteria, to lessen the chance of preexisting differences between the groups.

We will discuss here the two major types of quasi-experimental designs, as well as one type—ex post facto (after the fact) control group design—that is often mistakenly termed *quasi-experimental* (other types can be found in Cook & Campbell 1979; Mohr 1992):

- *Nonequivalent control group designs*—**Nonequivalent control group designs** have experimental and comparison groups that are designated before the treatment occurs but are not created by random assignment.
- *Before-and-after designs*—**Before-and-after designs** have a pretest and posttest but no comparison group. In other words, the subjects exposed to the treatment serve, at an earlier time, as their own control group. To

qualify as a quasi-experimental design, there must be more than one group with a before-and-after comparison on the same variable.

- *Ex post facto control group designs*—**Ex post facto control group designs** use nonrandomized control groups designated after the fact.

Exhibit 6.4 diagrams two studies, one using a nonequivalent control group design and another using the multiple group before-and-after design; the two studies are discussed subsequently. (An ex post facto control group design is the same as for a nonequivalent control group design, but the two types of experiment differ in how people are able to join the groups.)

Journal Link

Read about a quasi-experimental design that links music training and memory.

Quasi-experiments can establish an association of variables: How well do they meet the other criteria for showing causal relationships? If quasi-experimental designs are longitudinal, they can establish time order. But these designs are weaker than true experiments in establishing nonspuriousness: They aren't good at ruling out the influence of some third, uncontrolled variable. Because quasi-experiments do not require random assignment, they can be conducted using more natural procedures in more natural settings, so we may gain a more complete understanding of causal context. However, quasi-experiments are neither better nor worse than experiments in identifying the mechanism of a causal effect.

Nonequivalent Control Group Designs

In this type of quasi-experimental design, a comparison group is selected to be as comparable as possible to the treatment group. Two selection methods can be used:

1. *Individual matching*—Individual cases in the treatment group are matched with similar individuals in the comparison group. This can sometimes create a comparison group that is very similar to the experimental group, such as when Head Start participants were matched with their siblings to estimate the effect of participation in Head Start. However, in many studies, it may not be possible to match on the most important variables.
2. *Aggregate matching*—In most situations when random assignment is not possible, the second method of matching makes more sense: identifying a comparison group that matches the treatment group in the aggregate rather than trying to match individual cases. This means finding a comparison group that has similar distributions on key variables: the same average age, the same percentage female, and so on. The upper part of Exhibit 6.4 diagrams a study done at Xerox Corporation by Ruth Wageman (1995), in which 152 technical service teams were divided into three experimental conditions. One emphasized a group orientation with interdependent tasks; another emphasized a "hybrid" style, with some interdependent and some individual tasks; the third group of teams worked as individual technicians. All were evaluated before and after on their performance. The groups were roughly—though not vigorously—equivalent before the study; their leaders chose which style they would pursue, so the procedure was not a true experiment. Interestingly, the hybrid condition proved less successful than either the group or individual approach.

Exhibit 6.4 Quasi-Experimental Designs

Nonequivalent control group design: Interdependence and team performance (Wageman 1995)				
Experimental group:		O_1	X_a	O_2
Comparison group 1:		O_1	X_b	O_2
Comparison group 2:		O_1	X_c	O_2
		Pretest	*Treatment*	*Posttest*
Team interdependence	Group	Team performance	Interdependent tasks	Team performance
	Hybrid	Team performance	Mixed tasks	Team performance
	Individual	Team performance	Individual tasks	Team performance
Before-and-after design: Soap opera suicide and actual suicide (Phillips 1982)				
Experimental group:		O_{11}	X_1	O_{21}
		O_{12}	X_2	O_{22}
		O_{13}	X_3	O_{23}
		O_{14}	X_4	O_{24}
		Pretest	*Treatment*	*Posttest*
		Suicide rate	Soap opera suicides	Suicide rate
Key: O = Observation (pretest or posttest) X = Experimental treatment				

Source: Wageman, Ruth. 1995. Interdependence and group effectiveness. Administrative Science Quarterly 40: 145–180. Published by Sage Publications on behalf of Johnson Graduate School of Management, Cornell University.

Nonequivalent control group designs allow you to determine whether an association exists between the presumed cause and effect.

Quasi-experimental design: A research design in which there is a comparison group that is comparable to the experimental group in critical ways but subjects are not randomly assigned to the comparison and experimental groups.

Nonequivalent control group design: A quasi-experimental design in which there are experimental and comparison groups that are designated before the treatment occurs but are not created by random assignment.

Before-and-after design: A quasi-experimental design consisting of several before-and-after treatment comparisons involving the same variables but no comparison group.

Ex post facto control group design: A nonexperimental design in which comparison groups are selected after the treatment, program, or other variation in the independent variable has occurred.

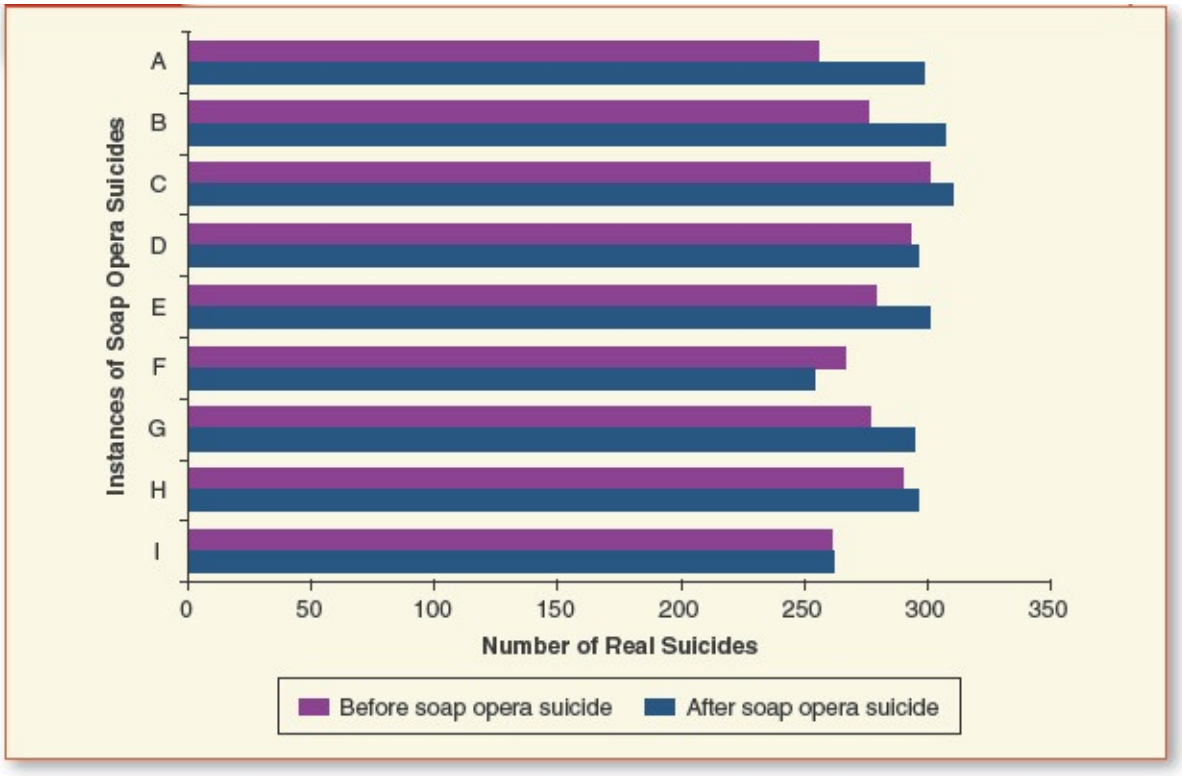
Before-and-After Designs

The common feature of before-and-after designs is the absence of a comparison group: All cases are exposed to the experimental treatment. The basis for comparison is instead provided by the pretreatment measures in the experimental group. These designs are thus useful for studies of interventions that are experienced by virtually every case in some population, such as total coverage programs such as Social Security or single-organization studies of the effect of a new management strategy.

The simplest type of before-and-after design is the fixed-sample panel design. As you may recall from Chapter 2, in a panel design, the same individuals are studied over time; the research may entail one pretest and one posttest. However, this type of before-and-after design does not qualify as a quasi-experimental design because comparing subjects to themselves at just one earlier point in time does not provide an adequate comparison group. Many influences other than the experimental treatment may affect a subject following the pretest—for instance, basic life experiences for a young subject.

A more powerful, **multiple group before-and-after design** is illustrated by David P. Phillips's (1982) study of the effect of TV soap opera suicides on the number of actual suicides in the United States. In this study, before-and-after comparisons were made of the same variables between different groups, as illustrated in the bottom half of Exhibit 6.4. Phillips identified 13 (fictional) soap opera suicides in 1977 and then recorded the actual U.S. suicide rate in the weeks before and following each TV story. In effect, the researcher had 13 different before-and-after studies, 1 for each suicide story. In 12 of these 13 comparisons, real deaths from suicide increased from the week before each soap opera suicide to the week after (Exhibit 6.5). Phillips also found similar increases in motor vehicle deaths and crashes during the same period, some portion of which reflects covert suicide attempts.

Exhibit 6.5 Real Suicides and Soap Opera Suicides



Source: Phillips, David P. 1982. The impact of fictional television stories on U.S. adult fatalities: New evidence on the effect of the mass media on violence. American Journal of Sociology 87(May): 1347. Reprinted with permission from the University of Chicago Press.

Another type of before-and-after design involves multiple pretest and posttest observations of the same group. **Repeated measures panel designs** include several pretest and posttest observations, allowing the researcher to study the process by which an intervention or treatment has an impact over time; hence, they are better than simple before-and-after studies.

Time series designs include many (preferably 30 or more) such observations in both pretest and posttest periods. They are particularly useful for studying the impact of new laws or social programs that affect large numbers of people and that are readily assessed by some ongoing measurement. For example, we might use a time series design to study the impact of a new seat belt law on the severity of injuries in automobile accidents, using a monthly state government report on insurance claims. Special statistics are required to analyze time series data, but the basic idea is simple: Identify a trend in the dependent variable up to the date

of the intervention, and then project the trend into the postintervention period. This *projected* trend is then compared with the *actual* trend of the dependent variable after the intervention. A substantial disparity between the actual and projected trends is evidence that the intervention or event had an impact (Rossi & Freeman 1989: 260–261, 358–363).

How well do these before-and-after designs meet the five criteria for establishing causality? The before-and-after comparison enables us to determine whether an *association* exists between the intervention and the dependent variable (because we can determine whether a change occurred after the intervention). They also clarify whether the change in the dependent variable occurred after the intervention, so *time order* is not a problem. However, there is no control group, so we cannot rule out the influence of extraneous factors as the actual cause of the change we observed; *spuriousness* may be a problem.

Overall, the longitudinal nature of before-and-after designs can help identify causal *mechanisms*, while the loosening of randomization requirements makes it easier to conduct studies in natural settings, where we learn about the influence of *contextual* factors.

Multiple group before-and-after design: A type of quasi-experimental design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.

Repeated measures panel design: A quasi-experimental design consisting of several pretest and posttest observations of the same group.

Time series design: A quasi-experimental design consisting of many pretest and posttest observations of the same group.

Ex Post Facto Control Group Designs

Groups in ex post facto designs are designated after the treatment has occurred; hence, ex post facto studies fail even to earn the quasi-experimental designation. The problem is that people were neither randomly assigned, nor were they picked for experimental treatments. They may well have selected themselves into (or out of) treatment groups. Of course, this makes it difficult to determine whether an association between group membership and outcome is spurious. However, the particulars will vary from study to study; in some circumstances, we may conclude that the treatment and control groups are so similar that causal effects can be tested (Rossi & Freeman 1989: 343–344).

Susan Cohen and Gerald Ledford Jr.'s (1994) study of the effectiveness of self-managing teams used a well-constructed ex post facto design. They studied a telecommunications company with some work teams that were self-managing and some that were traditionally managed (meaning that a manager was responsible for the team's decisions). Cohen and Ledford found the self-reported quality of work life to be higher in the self-managing groups than in the traditionally managed groups.

What Are the Threats to Validity?

Any research design should be evaluated for its ability to yield valid conclusions, and different designs have different strengths in this regard. Remember, there are three kinds of validity: (1) internal (or causal), (2) external (or generalizability), and (3) measurement. True experiments are good at producing internal validity, but fare less well in achieving external validity (generalizability). Quasi-experiments, by comparison, may provide more generalizable results than true experiments but are more prone to problems of internal invalidity. Nonexperimental designs such as those used in survey or field research are often weaker at internal validity but stronger on generalizability, with no particular advantage in measurement validity. In this section, we describe a host of problems that arise, in experiments and other methods, with establishing internal validity and generalizability. These are perennial, persistent problems in social research of all kinds.

Interactive Exercises Link

Souces of Internal Validity

Threats to Internal (Causal) Validity

The following sections discuss 11 *threats to validity* (sometimes referred to as *sources of invalidity*) that occur frequently in social science research (Exhibit 6.6). These “threats” exemplify five major types of problems that arise in research design.

Video Link

Watch a lecture on threats to internal validity.

Noncomparable Groups

The problem of noncomparable groups occurs when the experimental group and the control group are not really comparable—that is, when something interferes with the two groups being essentially the same at the beginning (or end) of a study.

Exhibit 6.6 Threats to Internal Validity

Problem	Example	Type
Selection	Girls who choose to see a therapist are not representative of population.	Noncomparable Groups
Mortality	Students who most dislike college drop out, so aren't surveyed.	Noncomparable Groups
Instrument Decay	Interviewer tires, losing interest in later interviews, so poor answers result.	Noncomparable Groups
Testing	If someone has taken the SAT before, they are familiar with the format, so do better.	Endogenous Change
Maturation	Everyone gets older in high school; it's not the school's doing.	Endogenous Change
Regression	The lowest-ranking students on IQ must improve their rank; they can't do worse.	Endogenous Change
History	Boston Marathon bombing affects marketing study of northeastern cities.	History
Contamination	"John Henry" effect; people in study compete with one another.	Contamination
Experimenter Expectation	Researchers unconsciously help their subjects, distorting results.	Treatment Misidentification
Placebo Effect	Fake pills in medical studies produce improved health.	Treatment Misidentification
Hawthorne Effect	Workers enjoy being subjects and work harder.	Treatment Misidentification

- *Selection bias* —When the subjects in your groups are initially different, **selection bias** occurs. If the ambitious students decide to be in the "coffee" group, you'll think their performance was helped by coffee—but it could have been their ambition.

Examples of selection bias are everywhere. Harvard graduates are very successful people, but Harvard *admits* students who are likely to be successful anyway. Maybe Harvard itself had no effect on them. A few years ago, a psychotherapist named Mary Pipher wrote a best seller called *Reviving Ophelia* (1994) in which she described the difficult lives—as she saw it—of typical adolescent girls. Pipher painted a stark picture of depression, rampant eating disorders, low self-esteem, academic failure, suicidal thoughts, and even suicide itself. Where did she get this picture? From her patients—that is, from adolescent girls who were in deep despair, or at least were unhappy enough to seek help. If Pipher had talked with a comparison sample of girls who hadn't

sought help, perhaps the story would not have been so bleak.

In the Sherman and Berk (1984) domestic violence experiment in Minneapolis, some police officers sometimes violated the random assignment plan when they thought the circumstances warranted arresting a suspect who had been randomly assigned to receive just a warning; thus, they created a selection bias in the experimental group.

- *Mortality*—Even when random assignment works as planned, the groups can become different over time because of **differential attrition**, or **mortality**; this can also be called *deselection*. That is, the groups become different because subjects in one group are more likely to drop out for various reasons compared with subjects in the other group(s). At some colleges, satisfaction surveys show that seniors are more likely to rate their colleges positively than are freshmen. But remember that the freshmen who really hated the place may have transferred out, so their ratings aren't included with senior ratings. In effect, the lowest scores are removed; that's a mortality problem. This is not a likely problem in a laboratory experiment that occurs in one session, but some laboratory experiments occur over time, so differential attrition can become a problem. Subjects who experience the experimental condition may become more motivated to continue in the experiment than are comparison subjects.

Note that whenever subjects are not assigned randomly to treatment and comparison groups, the threat of selection bias or mortality is great. Even if the comparison group matches the treatment group on important variables, there is no guarantee that the groups were similar initially for either the dependent variable or some other characteristic. However, a pretest helps the researchers to determine and control for selection bias.

- *Instrument decay*—Measurement instruments of all sorts wear out, in a process known as **instrument decay**, producing different results for cases studied later in the research. An ordinary spring-operated bathroom scale, for instance, becomes "soggy" after some years, showing slightly heavier weights than would be correct. Or a college teacher—a kind of instrument for measuring student performance—gets tired after reading too many papers one weekend and starts giving everyone a B. Research interviewers can get tired or bored, too, leading perhaps to shorter or less thoughtful answers from subjects. In all these cases, the measurement instrument has

"decayed," or worn out.

Selection bias: A source of internal (causal) invalidity that occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.

Differential attrition (mortality): A problem that occurs in experiments when comparison groups become different because subjects in one group are more likely to drop out for various reasons compared with subjects in the other group(s).

Instrument decay: The deterioration over time of a measurement instrument, resulting in increasingly inaccurate results.

Endogenous Change

The next three problems, subsumed under the label **endogenous change**, occur when natural developments in the subjects, independent of the experimental treatment itself, account for some or all of the observed change between pretest and posttest.

- *Testing*—Taking the pretest can itself influence posttest scores. As the Kaplan SAT prep courses attest, there is some benefit just to getting used to the test format. Having taken the test beforehand can be an advantage. Subjects may learn something or may be sensitized to an issue by the pretest and, as a result, respond differently the next time they are asked the same questions on the posttest.
- *Maturation*—Changes in outcome scores during experiments that involve a lengthy treatment period may be caused by maturation. Subjects may age, gain experience, or grow in knowledge—all as part of a natural maturational experience—and thus respond differently on the posttest than on the pretest. In many high school yearbooks, seniors are quoted as saying, for instance, "I started at West Geneva High as a boy and leave as a man. WGHS made me grow up." Well, he probably would have grown up anyway, high school or not. WGHS wasn't the cause.
- *Regression*—Subjects who are chosen for a study because they received very low scores on a test may show improvement in the posttest, on average, simply because some of the low scorers were having a bad day. Whenever subjects are selected for study because of extreme scores (either very high or very low), the next time you take their scores, they will likely

> “regress,” or move toward the average. After all, in a normal (bell curve) distribution, that’s what the average *is:* the most likely score. For instance, suppose you give an IQ test to third graders and then pull out the bottom 20% of the class for special attention. The next time that group (the 20%) take the test, they’ll almost certainly do better—and not just because of testing practice. In effect, they *can’t* do worse—they were at the bottom already. On average, they must do better. A football team that goes 0–12 one season almost has to improve. A first-time novelist writes a wonderful book and gains worldwide acclaim and a host of prizes. The next book is not so good, and critics say, “The praise went to her head.” But it didn’t; she almost *couldn’t* have done better. Whenever you pick people for being on an extreme end of a scale, odds are that next time, they’ll be more average. This is called the **regression effect**.

Testing, maturation, and regression effects are generally not a problem in experiments that have a control group because they would affect the experimental group and the comparison group equally. However, these effects could explain any change over time in most before-and-after designs because these designs do not have a comparison group. Repeated measures, panel studies, and time series designs are better in this regard because they allow the researcher to trace the pattern of change or stability in the dependent variable up to and after the treatment. Ongoing effects of maturation and regression can thus be identified and taken into account.

Endogenous change: A source of causal invalidity that occurs when natural developments or changes in the subjects (independent of the experimental treatment itself) account for some or all of the observed change from the pretest to the posttest.

Regression effect: A source of causal invalidity that occurs when subjects chosen because of their extreme scores on a dependent variable become less extreme on a posttest as a result of mathematical necessity, rather than the treatment.

History

History, or external events during the experiment (things that happen outside the experiment), could change subjects’ outcome scores. Examples are newsworthy events that concern the focus of an experiment and major disasters to which subjects are exposed. If you were test marketing promotional materials for various northeastern U.S. cities in April 2013, the results could be seriously

affected by the enormous publicity around Boston Marathon bombings, and the subsequent “Boston Strong” response. Such a problem is referred to as a **history effect**—history during the experiment, that is. Also called ***effect of external events***, it is a particular concern in before-and-after designs.

Causal conclusions can be invalid in some true experiments because of the influence of external events. For example, in an experiment in which subjects go to a special location for the treatment, something at that location unrelated to the treatment could influence these subjects. External events are a major concern in studies that compare the effects of programs in different cities or states (Hunt 1985: 276–277).

History effect (effect of external events): Events external to the study that influence posttest scores, resulting in causal invalidity.

Contamination: A source of causal invalidity that occurs when either the experimental or the comparison group is aware of the other group and is influenced in the posttest as a result.

Compensatory rivalry (John Henry effect): A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied the treatment and modify their efforts by way of compensation.

Demoralization: A type of contamination in experimental and quasi-experimental designs that occurs when control group members feel that they have been left out of some valuable treatment, performing worse than expected as a result.

Contamination

Contamination occurs in an experiment when the comparison and treatment groups somehow affect each other. When comparison group members know they are being compared, they may increase their efforts just to be more competitive. This has been termed **compensatory rivalry**, or the **John Henry effect**, named after the “steel-driving man” of the folk song, who raced against a steam drill in driving railroad spikes and killed himself in the process. Knowing that they are being denied some advantage, comparison group subjects may as a result increase their efforts to compensate. Conversely, comparison group members may experience **demoralization** if they feel that they have been left out of some valuable treatment, performing worse than expected as a result. Both compensatory rivalry and demoralization thus distort the impact of the experimental treatment.

The danger of contamination can be minimized if the experiment is conducted in a laboratory, if members of the experimental group and the comparison group have no contact while the study is in progress, and if the treatment is relatively brief. Whenever these conditions are not met, the likelihood of contamination increases.

Treatment Misidentification

Sometimes the subjects experience a “treatment” that wasn’t intended by the researcher. The following are three possible sources of **treatment misidentification**:

Research|Social Impact Link

Read about the power of the placebo effect.

1. *Expectancies of experiment staff*—Change among experimental subjects may result from the positive expectancies of experiment staff who are delivering the treatment rather than to the treatment itself. Even well-trained staff may convey their enthusiasm for an experimental program to the subjects in subtle ways. This is a special concern in evaluation research, when program staff and researchers may be biased in favor of the program for which they work and are eager to believe that their work is helping clients. Such positive staff expectations, the **expectancies of experiment staff**, thus create a **self-fulfilling prophecy**. However, in experiments on the effects of treatments such as medical drugs, **double-blind procedures** can be used: Staff delivering the treatments do not know which subjects are getting the treatment and which are receiving a placebo—something that looks like the treatment but has no intrinsic effect.
2. *Placebo effect*—In medicine, a *placebo* is a chemically inert substance (a sugar pill, for instance) that looks like a drug but actually has no direct physical effect. Research shows that such a pill can actually produce positive health effects in two thirds of patients suffering from relatively mild medical problems (Goleman 1993: C3). In other words, if you wish that a pill will help, it often actually does. In social science research, such **placebo effects** occur when subjects think their behavior should improve through an experimental treatment and then it does—not from the

treatment, but from their own beliefs. Researchers might then misidentify the treatment as having produced the effect.

3. *Hawthorne effect*—Members of the treatment group may change relative to the dependent variable because their participation in the study makes them feel special. This problem could occur when treatment group members compare their situation with that of members of the control group who are not receiving the treatment, in which case it would be a type of contamination effect. But experimental group members could feel special simply because they are in the experiment. This is termed a **Hawthorne effect** after a classic worker productivity experiment conducted at the Hawthorne electric plant outside Chicago in the 1920s. No matter what conditions the researchers changed to improve or diminish productivity (for instance, increasing or decreasing the lighting in the plant), the workers seemed to work harder simply because they were part of a special experiment. Oddly enough, some later scholars suggested that in the original Hawthorne studies, there was actually a selection bias, not a true Hawthorne effect—but the term has stuck (see Bramel & Friend 1981). Hawthorne effects are also a concern in evaluation research, particularly when program clients know that the research findings may affect the chances for further program funding.

Treatment misidentifications can sometimes be avoided through a technique called **process analysis** (Hunt 1985: 272–274). Periodic measures are taken throughout an experiment to assess whether the treatment is being delivered as planned. For example, Robert Drake and his colleagues (1996) collected process data to monitor the implementation of two employment service models that they tested. One site did a poorer job of implementing the individual placement and support model than the other site did, although the required differences between the experimental conditions were still achieved. Process analysis is often a special focus in evaluation research because of the possibility of improper implementation of the experimental program.

Treatment misidentification: A problem that occurs in an experiment when not the treatment itself, but rather some unknown or unidentified intervening process, is causing the outcome.

Expectancies of experiment staff (self-fulfilling prophecy): A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects results from the positive expectancies of the staff who are delivering the treatment, rather than to the treatment itself.

Double-blind procedure: An experimental method in which neither the subjects nor the staff

delivering experimental treatments know which subjects are getting the treatment.

Placebo effect: A source of treatment misidentification that can occur when subjects receive a treatment that they consider likely to be beneficial and improve as a result of that expectation rather than the treatment itself.

Hawthorne effect: A type of contamination in experimental and quasi-experimental designs that occurs when members of the treatment group change relative to the dependent variable because their participation in the study makes them feel special.

Process analysis: A research design in which periodic measures are taken to determine whether a treatment is being delivered as planned, usually in a field experiment.

Threats to Generalizability

Even true experimental designs have one major weakness, an Achilles' heel: The design components essential for true experiments that minimize threats to causal validity simultaneously make it more difficult to achieve both sample generalizability—being able to apply the findings to some clearly defined larger population—and cross-population generalizability—generalizing across subgroups and to other populations and settings.

Sample Generalizability

Subjects who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the duration of the study are unlikely to be a representative sample of any large population of interest to social scientists. Can they be expected to react to the experimental treatment in the same way as members of the larger population? The generalizability of the treatment and of the setting for the experiment also must be considered (Cook & Campbell 1979: 73–74): the more artificial the experimental arrangements, the greater the problem (Campbell & Stanley 1966: 20–21).

Audio Link

Listen to how researchers are beginning to generalize research on concussions.

In some limited circumstances, a researcher may be able to sample subjects randomly for participation in an experiment and thus select a generalizable sample—one that is representative of the population from which it is selected. This approach is occasionally possible in **field experiments**. For example, some studies of the effects of income supports on the work behavior of poor persons have randomly sampled persons within particular states before randomly assigning them to experimental and comparison groups. Sherman and Berk's (1984) field experiment about the impact of arrest in actual domestic violence incidents (see Chapter 2) used a slightly different approach. In this study, all eligible cases were treated as subjects in the experiment during the data

collection periods. As a result, we can place a good deal of confidence in the generalizability of the results to the population of domestic violence arrest cases in Minneapolis at the time.

One especially powerful type of field experiment is an audit (or paired testing) study, in which matched pairs of individuals (called *testers*) approach various organizations to discover how different people—for instance whites versus blacks, or men versus women—are treated. Audit studies were developed and widely used in the 1970s first to uncover housing discrimination. More recently, they have been used in research on employment (Cross et al. 1990), automobile purchases (Ayres & Siegelman 1995), restaurant hiring (women have more difficulty being hired in expensive restaurants) (Neumark 1996), and even taxicab rides (Ayres, Vars, & Zakariya 2005). Audit researchers try to make testers as similar as possible in every respect but the one trait they wish to test for (e.g., race or gender).

What effect, for example, might a criminal record noted on one's job application have on a man's chance of getting a job? A huge effect, as it happens—reducing the chance of getting a callback after submitting an application by at least 50%. In a study of 350 employers in Milwaukee, Wisconsin, Devah Pager (2003) used pairs of white and black testers, rotating which testers claimed a criminal record. Pager found that a (supposed) criminal record reduced white men's chances of a callback by one half, and black men's chances by two thirds. And black men—apart from the criminal record—were already seriously discriminated against. All told, a white man *with* a criminal record was more likely to be called than a black man *without* a criminal record. In a follow-up study, Pager and Lincoln Quillian (2005) found that the same employers who said they didn't discriminate against black men, or even against a criminal record, in fact—when faced with a live applicant—did discriminate against both, very significantly. The audit study showed that a survey was a poor indicator of what employers actually did.

Researchers using audit studies have to be careful to match the testers well, to make sure that no unintended differences (e.g., speech patterns, clothing styles) exist that might affect the results and to train testers well so that they aren't inadvertently influencing people in the audited organizations, or seeing discrimination where there may be none. The generalizability of audit studies is also limited because of their focus on entry-level positions and, in employment studies, on "call back" outcomes (rather than, say, employment or salary offers) (Favreault 2008; Heckman & Siegelman 1993). And of course, the procedure

used to select the employers or other organizations also determines the generalizability of an audit study's findings.

Field experiment: An experimental study conducted in a real-world setting.

Cross-Population Generalizability

Researchers often are interested in determining whether treatment effects identified in an experiment hold true across different populations, times, or settings. When random selection is not feasible, the researchers may be able to increase the cross-population generalizability of their findings by selecting several different experimental sites that offer marked contrasts on key variables (Cook & Campbell 1979: 76–77).

Within a single experiment, researchers also may be concerned with whether the relationship between the treatment and the outcome variable holds true for certain subgroups. This demonstration of *external validity* is important evidence about the conditions that are required for the independent variable(s) to have an effect. Richard Price, Michelle Van Ryn, and Amiram Vinokur (1992) found that intensive job search assistance reduced depression among individuals who were at high risk for it because of other psychosocial characteristics; however, the intervention did not influence the rate of depression among individuals at low risk for depression. This is an important limitation on the generalizability of the findings, even if the sample Price and colleagues took was representative of the population of unemployed persons.

Finding that effects are consistent across subgroups does not establish that the relationship also holds true for these subgroups in the larger population, but it does provide supportive evidence. We have already seen examples of how the existence of treatment effects in particular subgroups of experimental subjects can help us predict the cross-population generalizability of the findings. For example, Sherman and Berk's research (1984; see Chapter 2) found that arrest did not deter subsequent domestic violence for unemployed individuals; arrest also failed to deter subsequent violence in communities with high levels of unemployment.

There is always an implicit trade-off in experimental design between maximizing causal validity, on the one hand, and generalizability, on the other.

Research subjects willing to be randomized into groups and experimented on are probably not representative of the larger population. College students, to take an important example, are easy to recruit and to assign to artificial but controlled manipulations, so they are frequently the subjects in experimental psychology research, but again, the generalizability to other groups may be uncertain. In a fascinating and clever series of experiments, Andrew Elliott and Daniela Nesta (2008) examined how the color red affected men's rating of a woman's attractiveness. They sorted male undergraduates randomly into two groups, then showed them head shots of a moderately attractive young woman, with the photograph bordered either by white (the control group) or by red (the treatment group). The woman in the red-framed picture was rated as significantly more attractive. The researchers then compared men with women raters, also looking at photos with differently colored frames; the female raters were unaffected by color. And, the ratings were found to be specifically on sexual attractiveness, not "likeability." In a series of studies, Elliott and Nesta tried different colors, controlled for sexual orientation, and ensured that subjects were not aware of the border color as a factor in their judgments. "Red," they found, "leads men to view women as more attractive and more sexually desirable" (p. 1150). The limitation may be that their research was on undergraduates; it may be that the "red" effect may not be generalizable or is less powerful, say, for older men—or, for that matter, older women who are being judged. From this research, we can't know.

Although we need to be skeptical about the generalizability of the results of a single experiment or setting, the body of findings accumulated from many experimental tests with different people in different settings can provide a solid basis for generalization (Campbell & Russo 1999: 143).

Interaction of Testing and Treatment

A variant on the problem of external validity occurs when the experimental treatment has an effect only when particular conditions created by the experiment occur. One such problem occurs when the treatment has an effect only if subjects have had the pretest. The pretest sensitizes the subjects to some issue so that when they are exposed to the treatment, they react in a way they would not have if they had not taken the pretest. In other words, testing and treatment interact to produce the outcome. For example, answering questions in a pretest about racial prejudice may sensitize subjects so that when they are

exposed to the experimental treatment, seeing a film about prejudice, their attitudes are different from what they would have been. In this situation, the treatment truly had an effect, but it would not have had an effect if it were repeated without the sensitizing pretest. This possibility can be evaluated by using the Solomon four-group design to compare groups with and without a pretest (Exhibit 6.7). If testing and treatment do interact, the difference in outcome scores between the experimental and comparison groups will be different for subjects who took the pretest and those who did not.

Exhibit 6.7 Solomon Four-Group Design Testing the Interaction of Pretesting and Treatment

Experimental group:	R	O_1	X	O_2
Comparison group:	R	O_1		O_2
Experimental group:	R		X	O_2
Comparison group:	R			O_2

Key: R = Random assignment
O = Observation (pretest or posttest)
X = Experimental treatment

As you can see, no single procedure establishes the external validity of experimental results. Ultimately, we must base our evaluation of external validity on the success of replications taking place at different times and places and using different forms of the treatment.

How Do Experimenters Protect Their Subjects?

Social science experiments often involve subject deception. Primarily because of this feature, some experiments have prompted contentious debates about research ethics. Experimental evaluations of social programs also pose ethical dilemmas because they require researchers to withhold possibly beneficial treatment from some of the subjects just on the basis of chance. Such research may also yield sensitive information about program compliance, personal habits, and even illegal activity—information that is protected from legal subpoenas only in some research concerning mental illness or criminal activity (Boruch 1997). In this section, we give special attention to the problems of deception and the distribution of benefits in experimental research.

Deception

Deception occurs when subjects are misled about research procedures to determine how they would react to the treatment if they were not research subjects. Deception is a critical component of many social experiments, partly because of the difficulty of simulating real-world stresses and dilemmas in a laboratory setting. Stanley Milgram's (1963) classic study of obedience to authority provides a good example. (If you have read Chapter 3 already, you'll be familiar with this example.) Volunteers were recruited for what they were told was a study of the learning process. The experimenter told the volunteers they were to play the role of "teacher" and to administer an electric shock to a "student" in the next room when the student failed a memory test. The shocks were phony (and the students were actors), but the real subjects, the volunteers, didn't know this. They were told to increase the intensity of the shocks, even beyond what they were told was a lethal level. Many subjects continued to obey the authority in the study (the experimenter), even when their obedience involved administering what they thought were potentially lethal shocks to another person.

> But did the experimental subjects actually believe that they were harming someone? Observational data suggest they did: "Persons were observed to sweat, tremble, stutter, bite their lips, and groan as they found themselves increasingly implicated in the experimental conflict. (Milgram 1965: 66)

Verbatim transcripts of the sessions also indicated that participants were in much psychological agony about administering the "shocks." So it seems that Milgram's deception "worked." Moreover, it seemed "necessary" because Milgram could not have administered real electric shocks to the students, nor would it have made sense for him to order the students to do something that wasn't so troubling, nor could he have explained what he was really interested in before conducting the experiment. Here is the real question: Is this sufficient justification to allow the use of deception?

Elliot Aronson and Judson Mills's study (1959) of severity of initiation (at an all-women's college in the 1950s), also mentioned in Chapter 3, provides a very different example of the use of deception in experimental research—one that

does not pose greater-than-everyday risks to subjects. The students who were randomly assigned to the "severe initiation" experimental condition had to read a list of embarrassing words. Even in the 1950s, reading a list of potentially embarrassing words in a laboratory setting and listening to a taped discussion were unlikely to increase the risks to which students were exposed in their everyday lives. Moreover, the researchers informed subjects that they would be expected to talk about sex and could decline to participate in the experiment if this requirement would bother them. No one dropped out.

To further ensure that no psychological harm was caused, Aronson and Mills (1959) explained the true nature of the experiment to the subjects after the experiment, in what is called debriefing, also discussed in Chapter 3. The subjects' reactions were typical: "None of the Ss expressed any resentment or annoyance at having been misled. In fact, the majority were intrigued by the experiment, and several returned at the end of the academic quarter to ascertain the result" (p. 179). Although the American Sociological Association's (1997) *Code of Ethics* does not discuss experimentation explicitly, one of its principles highlights the ethical dilemma deceptive research poses:

> (a) Sociologists do not use deceptive techniques (1) unless they have determined that their use will not be harmful to research participants; is justified by the study's prospective scientific, educational, or applied value; and that equally effective alternative procedures that do not use deception are not feasible, and (2) unless they have obtained the approval of institutional review boards or, in the absence of such boards, with another authoritative body with expertise on the ethics of research.
>
> (b) Sociologists never deceive research participants about significant aspects of the research that would affect their willingness to participate, such as physical risks, discomfort, or unpleasant emotional experiences. (p. 3)

Selective Distribution of Benefits

Field experiments conducted to evaluate social programs also can involve issues of informed consent (Hunt 1985: 275–276). One ethical issue that is somewhat unique to field experiments is the **distribution of benefits**: How much are subjects harmed by the way treatments are distributed in the experiment? For example, Sherman and Berk's (1984) experiment, and its successors, required police to make arrests in domestic violence cases largely on the basis of a random process. When arrests were not made, did the subjects' abused spouses suffer? Price and colleagues (1992) randomly assigned unemployed individuals who had volunteered for job-search help to an intensive program. Were the unemployed volunteers who were assigned to the comparison group at a big disadvantage?

Is it ethical to give some potentially advantageous or disadvantageous treatment to people on a random basis? Random distribution of benefits is justified when the researchers do not know whether some treatment actually is beneficial or not —and, of course, it is the goal of the experiment to find out. Chance is as reasonable a basis for distributing the treatment as any other. Also, if insufficient resources are available to fund fully a benefit for every eligible person, distribution of the benefit on the basis of chance to equally needy persons is ethically defensible (Boruch 1997: 66–67).

Distribution of benefits: An ethical issue about how much researchers can influence the benefits subjects receive as part of the treatment being studied in a field experiment.

Conclusion

Causal (internal) validity is the last of the three legs on which the validity of research rests (the first two being valid measurement and generalizability). In this chapter, you have learned about the five criteria used to evaluate the causal validity of particular research designs. You have seen the problem of spuriousness and the way that randomization deals with it.

True experiments help greatly to achieve more valid causal conclusions—they are the “gold standard” for testing causal hypotheses. But even when conditions preclude a true experiment, adding experimental components can improve many research designs. However, although it may be possible to test a hypothesis with an experiment, it is not always desirable to do so. Laboratory experiments may be inadvisable when they do not test the real hypothesis of interest but test instead a limited version that is amenable to laboratory manipulation. It also may not make sense to test the impact of social programs that cannot actually be implemented because of financial or political problems (Rossi & Freeman 1989: 304–307). Yet the virtues of experimental designs mean that they should always be considered when explanatory research is planned.

Understandings of causal relationships are always partial. Researchers must always wonder whether they have omitted some relevant variables from their controls or whether their experimental results would differ if the experiment were conducted in another setting or at another time in history. But the tentative nature of causal conclusions means that we must give more—not less—attention to evaluating the causal validity of social science research whenever we need to ask the simple question, what caused variation in this social phenomenon?

Key Terms

Association 112
Before-and-after design 120
Causal effect 112
Ceteris paribus 112
Comparison groups 116
Compensatory rivalry (John Henry effect) 127
Contamination 127
Context 115
Control group 116
Demoralization 127
Differential attrition (mortality) 125
Distribution of benefits 132
Double-blind procedure 127
Endogenous change 126
Expectancies of experiment staff (self-fulfilling prophecy) 127
Experimental group 116
Ex post facto control group design 120
Field experiment 128
Hawthorne effect 128
History effect (effect of external events) 126
Instrument decay 125
Matching 117
Mechanism 115
Multiple group before-and-after design 122
Nonequivalent control group design 120
Nonspuriousness 113
Placebo effect 127
Posttest 117
Pretest 117
Process analysis 128
Quasi-experimental design 120
Random assignment (randomization) 117
Regression effect 126
Repeated measures panel design 123
Selection bias 125

Spurious 113
Time order 113
Time series design 123
Treatment misidentification 127
True experiment 116

Highlights

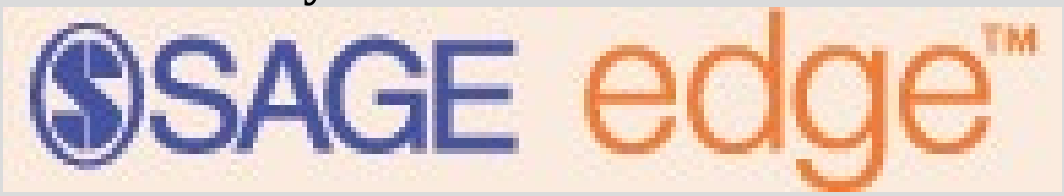
- Three criteria generally are viewed as necessary for identifying a causal relationship: association between the variables, proper time order, and nonspuriousness of the association. In addition, identification of a causal mechanism and the context strengthens the basis for concluding that a causal relationship exists.
- Association between two variables by itself is insufficient evidence of a causal relationship. This point is commonly made by the expression, “Correlation does not prove causation.”
- The independent variable in an experiment is represented by a treatment or other intervention. Some subjects receive one type of treatment; others may receive a different treatment or no treatment. In true experiments, subjects are assigned randomly to comparison groups.
- Experimental research designs have three essential components: use of at least two groups of subjects for comparison, measurement of the change that occurs as a result of the experimental treatment, and use of random assignment. In addition, experiments may include identification of a causal mechanism and control over experimental conditions.
- Random assignment of subjects to experimental and comparison groups eliminates systematic bias in group assignment. The odds of there being a difference between the experimental and comparison groups on the basis of chance can be calculated. They become very small for experiments with at least 30 subjects per group. Random assignment and random sampling both rely on a chance selection procedure, but their purposes differ. Random assignment involves placing predesignated subjects into two or more groups on the basis of chance; random sampling involves selecting subjects from a larger population on the basis of chance. Matching of cases in the experimental and comparison groups is a poor substitute for randomization, because identifying in advance all important variables on which to make the match is not possible. However, matching can improve the comparability of groups when it is used to supplement randomization.
- Ethical and practical constraints often preclude the use of experimental designs.
- A quasi-experimental design can be either a nonequivalent control group design or a before-and-after design. Nonequivalent control groups can be created through either individual matching of subjects or matching of group

characteristics. In either case, these designs can allow us to establish the existence of an association and the time order of effects, but they do not ensure that some unidentified extraneous variable did not cause what we think of as the effect of the independent variable. Before-and-after designs can involve one or more pretests and posttests. Although multiple pretests and posttests make it unlikely that another, extraneous influence caused the experimental effect, they do not guarantee it.

- Ex post facto control group designs include a comparison group that individuals could have decided to join precisely because they prefer this experience rather than what the experimental group offers. This creates differences in subject characteristics between the experimental and control groups, which might very well result in a difference in the dependent variable. Because of this possibility, this type of design is not considered a quasi-experimental design.
- Causal conclusions derived from experiments can be invalid because of selection bias, endogenous change, the history effects (effects of external events), cross-group contamination, or treatment misidentification. In true experiments, randomization should eliminate selection bias and bias resulting form endogenous change. External events, cross-group contamination, and treatment misidentification can threaten the validity of causal conclusions in both true experiments and quasi-experiments.
- Process analysis can be used in experiments to identify how the treatment had (or didn't have) an effect—a matter of particular concern in field experiments. Treatment misidentification is less likely when process analysis is used.
- The generalizability of experimental results declines if the study conditions are artificial and the experimental subjects are unique. Field experiments are likely to produce more generalizable results than experiments conducted in the laboratory.
- The external validity of causal conclusions is determined by the extent to which they apply to different types of individuals and settings. When causal conclusions do not apply to all the subgroups in a study, they are not generalizable to corresponding subgroups in the population; consequently, they are not externally valid with respect to those subgroups. Causal conclusions can also be considered externally invalid when they occur only under the experimental conditions.
- Subject deception is common in laboratory experiments and poses unique ethical issues. Researchers must weigh the potential harm to subjects and debrief subjects who have been deceived. In field experiments, a common

ethical problem is selective distribution of benefits. Random assignment may be the fairest way of allocating treatment when treatment openings are insufficient for all eligible individuals and when the efficacy of the treatment is unknown.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. There's a lot of "sound and fury" in the social science literature about units of analysis and levels of explanation. Some social researchers may call another a "reductionist" if the researcher explains a problem, such as substance abuse, as resulting from "lack of self-control." The idea is that the behavior requires consideration of social structure—a group level of analysis rather than an individual level of analysis. Another researcher may be said to commit an "ecological fallacy" if she assumes that group-level characteristics explain behavior at the individual level (such as saying that "immigrants are more likely to commit crime" because the neighborhoods with higher proportions of immigrants have higher crime rates). Do you favor causal explanations at the individual or the group (or social structural) level? If you were forced to mark on a scale from 0 to 100 the percentage of crime that results from problems with individuals rather than from problems with the settings in which they live and other aspects of social structure, where would you make your mark? Explain your decision.
2. Researchers often try to figure out how people have changed over time by conducting a cross-sectional survey of people of different ages. The idea is that if people who are in their 60s tend to be happier than people who are in their 20s, it is because people tend to "become happier" as they age. But maybe people who are in their 60s now were just as happy when they were in their 20s and people in their 20s now will be just as unhappy when they are in their 60s. (That's called a cohort effect.) We can't be sure unless we conduct a panel study (survey the same people at different ages). What, in your experience, are the major differences between the generations today in social attitudes and behaviors? Which would you attribute to changes as people age and which to differences between cohorts in what they have experienced (such as common orientations among baby boomers)? Explain your reasoning.
3. The chapter begins with some alternative explanations for recent changes in the homicide rate. Which of the explanations make the most sense to you? Why? How could you learn more about the effect on crime of one of the "causes" you have identified in a laboratory experiment? What type of study could you conduct in the community to assess its causal impact?
4. This chapter discusses both experimental and quasi-experimental approaches to identifying causes. What are the advantages and disadvantages of both approaches for achieving each of the five criteria identified for causal explanations?

Finding Research

1. Read an original article describing a social experiment. (Social psychology *readers*, collections of such articles for undergraduates, are a good place to find interesting studies.) Critique the article, using as your guide the article review questions presented in Exhibit 13.2 on page 318. Focus on the extent to which experimental conditions were controlled and the causal mechanism was identified. Did inadequate control over conditions or inadequate identification of the causal mechanism make you feel uncertain about the causal conclusions?
2. Go to the website of the Community Policing Consortium (www.policing.com/links/index.html). What causal assertions are made? Pick one of these assertions and propose a research design with which to test this assertion. Be specific.
3. Go to Sociosite (www.sociosite.net/). Choose "Subject Areas," and pick a sociological subject area you are interested in. Find an example of research that has been done using experimental methods in this subject. Explain the experiment. Choose at least five of the Key Terms listed at the end of this chapter that are relevant to and incorporated in the research experiment you have located on the Internet. Explain how each of the five Key Terms you have chosen plays a role in the research example you found on the web.

Critiquing Research

1. From newspapers or magazines, find two recent studies of education (reading, testing, etc.). For each study, list in order what you see as the most likely sources of internal invalidity (selection, mortality, etc.).
2. Select a true experiment, perhaps from the *Journal of Experimental and Social Psychology*, the *Journal of Personality and Social Psychology*, or sources suggested in class. Diagram the experiment using the exhibits in this chapter as a model. Discuss the extent to which experimental conditions were controlled and the causal mechanism was identified. How confident can you be in the causal conclusions from the study, based on review of the threats to internal validity discussed in this chapter: selection bias, endogenous change, external events, contamination, and treatment misidentification? How generalizable do you think the study's results are to the population from which the cases were selected? To specific subgroups in the study? How thoroughly do the researchers discuss these issues?
3. Repeat the previous exercise with a quasi-experiment.
4. Critique the ethics of one of the experiments presented in this chapter or some other experiment you have read about. What specific rules do you think should guide researchers' decisions about subject deception and the selective distribution of benefits?

Doing Research

1. Try out the process of randomization. Go to the Researcher Randomizer website (www.randomizer.org). Now just type numbers into the randomizer for an experiment with two groups and 20 individuals per group. Repeat the process for an experiment with 4 groups and 10 individuals per group. Plot the numbers corresponding to each individual in each group. Does the distribution of numbers within each group truly seem to be random?
2. Participate in a social psychology experiment on the Internet at the Social Psychology Network website (www.socialpsychology.org/expts.htm). Pick an experiment in which to participate and follow the instructions. After you finish, write a description of the experiment and evaluate it using the criteria discussed in the chapter.
3. Volunteer for an experiment. Contact the psychology department at your school and ask about opportunities for participating in laboratory experiments. Discuss the experience with your classmates.

Ethics Questions

1. Randomization is a key feature of experimental designs that are often used to investigate the efficacy of new treatments for serious and often incurable terminal diseases. What ethical issues do these techniques raise in studies of experimental treatments for incurable, terminal diseases? Would you make an ethical argument that in some situations, it is more ethical to use random assignment than the usual procedures for deciding whether patients receive a new treatment?
2. In their study of “neighborhood effects” on crime, sociologists Robert Sampson and Stephen Raudenbush (1999) had observers drive down neighborhood streets in Chicago and record the level of disorder they observed. What should have been the observers’ response if they observed a crime in progress? What if they just suspected that a crime was going to occur? What if the crime was a drug dealer interacting with a driver at the curb? What if it was a prostitute soliciting a customer? What, if any, ethical obligation does a researcher studying a neighborhood have to residents in that neighborhood? Should research results be shared at a neighborhood forum?

Video Interview Questions

Listen to the researcher interview for Chapter 6 at edge.sagepub.com/chamblissmssw5e.

1. Why was it important for the research assistant to use a script in this study?
2. How did Professor Youngreen measure creative output in his study?

Chapter 7 Survey Research

©iStockphoto.com/wdstock

Learning Objectives

1. Explain the strengths and weaknesses of omnibus surveys.
2. Explain the problem of sampling on the dependent variable.
3. Discuss the advantages and disadvantages of including "don't know" and neutral responses among response choices and of using open-ended questions.
4. List the different methods for improving survey questions.
5. Outline a cover letter for a survey that contains each of the required elements.
6. List the strengths and weaknesses of each mode of survey design, giving particular attention to response rates.
7. Discuss the key ethical issues in survey research.

Some 6 months after the September 11, 2001, attacks on the World Trade Center and the Pentagon, a small group of students at Hamilton College and their professor, Dennis Gilbert (2002), conducted a nationwide survey of American Muslims. The survey found that nearly 75% of the respondents either knew someone who had, or had themselves, experienced anti-Muslim discrimination since the attacks. "You are demons," "Pig religion," "You guys did it," some were told. Respondents described actions such as "He spit in my face," "He pulled off my daughter's hijab [her head covering]"—the list of abuses went on. In all, 517 American Muslims were contacted, through a careful sampling procedure, and were interviewed via telephone by Gilbert's students and by employees of the Zogby International polling firm. This survey provided a snapshot of the views of an important segment of American society.

In this chapter, we will use the Muslim America project, a "youth and guns" survey also done by Gilbert, and other surveys to illustrate some key features of survey research. We explain the major steps in questionnaire design and then consider the features of four types of surveys, highlighting the unique problems attending each one and suggesting some possible solutions. (For instance, how do we develop an initial list—a sampling frame—of American Muslims?) We discuss ethics issues in the final section. By the chapter's end, you should be well on your way to becoming an informed consumer of survey reports and a knowledgeable developer of survey designs.

Why Is Survey Research So Popular?

Survey research collects information from a *sample of individuals* through their responses to *standardized questions.* As you probably have observed, a great many social scientists rely on surveys as their primary method of data collection. In fact, surveys have become so common that we cannot evaluate much of what we read in the newspaper or see on TV without having some understanding of this method of data collection (Converse 1984).

Survey research owes its popularity to three advantages: (1) versatility, (2) efficiency, and (3) generalizability. The *versatility* of surveys is apparent in the wide range of uses to which they are put, including opinion polls, election campaigns, marketing surveys, community needs assessments, and program evaluations. Surveys are *efficient* because they are a relatively fast means of collecting data on a wide range of issues at relatively little cost—ranging from about $10 to $15 per respondent in mailed surveys of the general population to $30 for a telephone survey and then as much as $300 for in-person interview surveys (F. J. Fowler, personal communication, January 7, 1998; see also Dillman 1982/1991; Groves & Kahn 1979/1991). Because they can be widely distributed to representative samples (see Chapter 5), surveys also help in achieving *generalizable* results.

Audio Link

Listen to what we can learn from survey research.

Perhaps the most efficient type of survey is an **omnibus survey**, which includes a range of topics of interest to different social scientists or to other sponsors. The General Social Survey (GSS) of the National Opinion Research Center at the University of Chicago is a prime example of an omnibus survey. It is a 90-minute interview administered biennially to a probability sample of almost 3,000 Americans, with a wide range of questions and topic areas chosen by a board of overseers. The resulting data sets are made available to many universities, instructors, and students (Davis & Smith 1992; National Opinion Research Center 1992).

Survey research: Research in which information is collected from a sample of individuals through their responses to a set of standardized questions.

Omnibus survey: A survey that covers a range of topics of interest to different social scientists.

Research in the News

What Can Surveys Uncover?

A survey of 2,000 retired New York City police officers uncovered some problems in crime reporting. The survey focused on manipulation of reported crimes, such as downgrading a crime to a less serious offense or discouraging individuals from filing reports. A subsample of recently retired officers (N = 871) showed that more than half had "personal knowledge" of such manipulation. Criminologist Eli Silverman is using this survey to shed light on the systemic culture of improper reporting.

1. What features of a survey do you think would make honest reporting of misbehavior like this more likely? Consider asking about the behaviors of acquaintances, the auspices of the survey, and the method of survey administration.
2. Based on your own experience, do you think surveys of college students can provide valid estimates of student grades? . . . of instructor performance? . . . of the use of alcohol or marijuana?

News source: Ruderman, Wendy. 2012. Crime report manipulation is common among New York police, study finds. *New York Times*, June 29: A17.

How Should We Write Survey Questions?

Questions are the centerpiece of survey research, so selecting good questions is the single most important concern for survey researchers. All hope for achieving measurement validity is lost unless the questions in a survey are clear and convey the intended meaning to respondents.

Question writing for a particular survey might begin with a brainstorming session or a review of previous surveys. The Muslim America survey began with students formulating questions with help from Muslim students and professors. Most professionally prepared surveys contain previously used questions as well as some new ones, but every question that is considered for inclusion must be reviewed carefully for clarity and for its ability to convey the intended meaning to the respondents.

Adherence to the following basic principles will go a long way toward ensuring clear and meaningful questions.

Be Clear; Avoid Confusing Phrasing

In most cases, a *simple, direct approach* to asking a question minimizes confusion ("Overall, do you enjoy living in Ohio?"). Use shorter rather than longer words and sentences: *brave* rather than *courageous; job concerns* rather than *work-related employment issues* (Dillman 2000: 52). Conversely, questions shouldn't be abbreviated so much that the results are ambiguous. The following simple statement is *too* simple:

Residential location: _____

Does it ask for town? Country? Street address? In contrast, asking, "In what city or town do you live?" focuses attention clearly on a specific geographic unit, a specific time, and a specific person.

Avoid *negative phrases or words*, especially **double negatives**: "Do you disagree that there should not be a tax increase?" Respondents have a hard time figuring out which response matches their sentiments. Such errors can easily be avoided with minor wording changes, but even experienced survey researchers can make this mistake.

Avoid **double-barreled questions**; these actually ask two questions but allow only one answer. For instance, "Our business uses reviews and incentive plans to drive employee behavior. Do you agree or disagree?" What if the business uses only reviews? How should respondents answer? Double-barreled questions can lead to dramatically misleading results. For example, during the Watergate scandal in the 1970s, the Gallup poll asked, "Do you think President Nixon should be impeached and compelled to leave the presidency, or not?" Only about a third of Americans said yes. But when the wording was changed to ask whether President Nixon should be brought to trial before the Senate, more than half answered yes. The first version combined impeachment—trial—with conviction and may have confused people (Kagay 1992: E5).

It is also important to identify clearly what kind of information each question is to obtain. Some questions focus on attitudes, or on what people say they want or how they feel. Some questions focus on beliefs, or what people think is true. Some questions focus on behavior, or on what people do. And some questions focus on attributes, or on what people are like or have experienced (Dillman

1978: 79–118; Gordon 1992). Rarely can a single question effectively address more than one of these dimensions at a time.

Double negative: A question or statement that contains two negatives, which can muddy the meaning of the question.

Double-barreled question: A single survey question that actually asks two questions but allows only one answer.

Minimize Bias

The words used in survey questions should not trigger biases, unless doing so is the researcher's conscious intent. Biased words and phrases tend to produce misleading answers. Some polls ask obviously loaded questions, such as "Isn't it time for Americans to stand up for morality and stop the shameless degradation of the airwaves?" Especially when describing abstract ideas (e.g., *freedom, justice, fairness*), your choice of words can dramatically affect how respondents answer. Take the difference between *welfare* and *assistance for the poor.* On average, surveys have found that public support for more assistance for the poor is about *39 percentage points higher* than for welfare (Smith 1987). Most people favor helping the poor; most people oppose welfare. The "truly needy" gain our sympathy, but "loafers and bums" do not.

Sometimes responses can be distorted through the lack of good alternative answers. For example, the Detroit Area Study (Turner & Martin 1984: 252) asked the following question: "People feel differently about making changes in the way our country is run. In order to keep America great, which of these statements do you think is best?" When the only two response choices were "We should be very cautious of making changes," or "We should be free to make changes," only 37% said that we should be free to make changes. However, when a stronger response choice was added suggesting that we should "constantly" make changes, 24% chose that response, and another 32% still chose the "free to make changes" response. So instead of 37%, we now had a total of 56% who seemed open to making changes in the way our country is run (Turner & Martin 1984: 252). Including the more extreme positive alternative (constantly make changes) made the less extreme positive alternative more attractive.

To minimize biased responses, researchers have to test reactions to the phrasing of a question.

Allow for Disagreement

Some respondents tend to “agree” with a statement just to avoid disagreeing. In a sense, they want to be helpful. You can see the impact of this human tendency in a 1974 Michigan Survey Research Center survey about crime and lawlessness in the United States (Schuman & Presser 1981). When one question stated that individuals were more to blame for crime than were social conditions, 60% of the respondents agreed. But when the question was rephrased so respondents were asked, “In general, do you believe that individuals or social conditions are more to blame for crime and lawlessness in the United States?” only 46% chose individuals.

As a rule, you should present both sides of attitude scales in the question itself (Dillman 2000: 61–62). The response choices themselves should be phrased to make each one seem as socially approved, as “agreeable,” as the others.

Most people, for instance, won’t openly admit to having committed a crime or other disreputable activities. In this situation, you should write questions that make agreement seem more acceptable. Rather than ask, “Have you ever shoplifted something from a store?” Dillman (2000) suggests “Have you ever taken anything from a store without paying for it?” (p. 25). Asking about a range of behaviors or attitudes can also facilitate agreeing with those that are socially unacceptable.

Don’t Ask Questions They Can’t Answer

Respondents should be *competent* to answer questions. Too many surveys expect accurate answers from people who couldn’t reasonably know the answers. One campus survey we’ve seen asked professors to agree or disagree with statements such as the following:

Video Link

Watch a webinar on creating good survey questions.

> “Minority students are made to feel they are second-class citizens.”
> “The Campus Center does a good job of meeting the informal needs of students.”
> “The Campus Center is where students go to meet one another and socialize informally.”
> “Alcohol contributes to casual sex among students.”

But of course, most professors are in no position to know the answers to these questions about students’ lives. To know what students do or feel, one should ask students, not professors. You should also realize that memory isn’t a perfect tool—most of us, for instance, cannot accurately report what we ate for lunch on a Tuesday 2 weeks ago. To get accurate lunch information, ask about today’s meal.

Sometimes your survey itself can sort people by competence so that they answer the appropriate questions. For instance, if you include a question about job satisfaction in a survey of the general population, first ask respondents whether they have a job. These **filter questions** create **skip patterns**. For example, respondents who answer no to one question are directed to skip ahead to another question, but respondents who answer yes go on to the **contingent question**. Skip patterns should be indicated clearly, as demonstrated in Exhibit 7.1.

Filter question: A survey question used to identify a subset of respondents who then are asked other questions.

Skip pattern: The unique combination of questions created in a survey by filter questions and contingent questions.

Contingent question: A question that is asked of only a subset of survey respondents.

Floaters: Survey respondents who provide an opinion on a topic in response to a closed-ended question that does not include a “Don’t know” option but who will choose “Don’t know” if it is available.

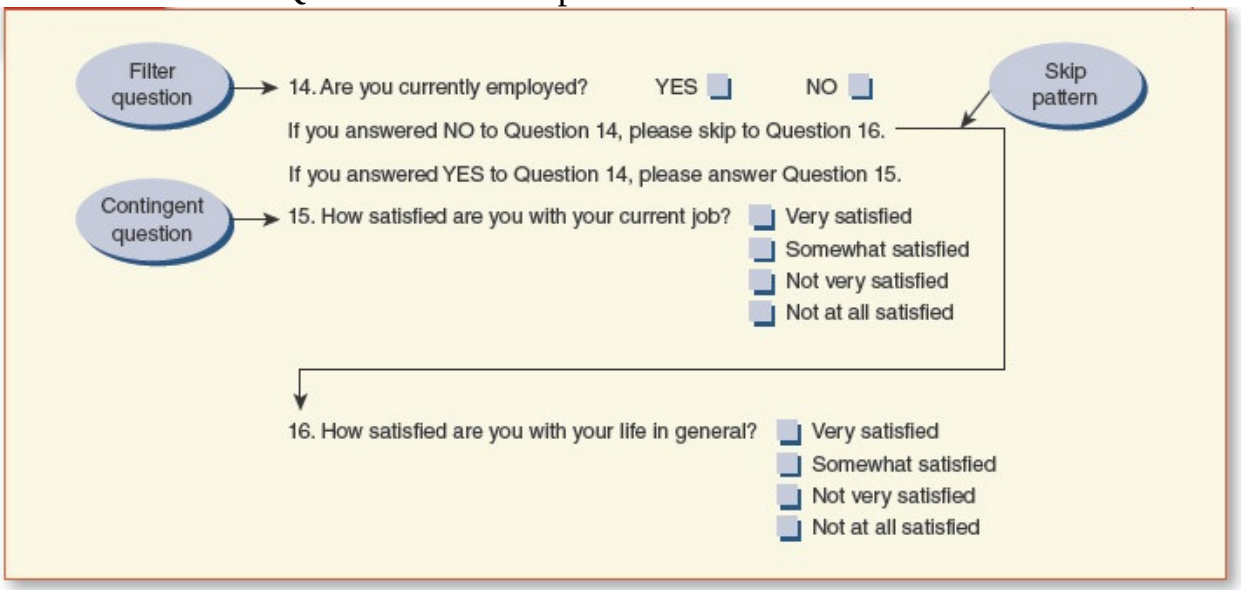
Allow for Uncertainty

Some respondents just don't know—about your topic, about their own feelings, about what they think. Or they like to be neutral and won't take a stand on anything. Or they don't have any information. All of these choices are okay, but you should recognize and allow for them.

Many people, for instance, are **floaters**: respondents who choose a substantive answer even when they really don't know. Asked for their opinion on a law of which they're completely ignorant, a third of the public will give an opinion anyway, if "Don't know" isn't an option. But if it *is* an option, 90% of that group will pick that answer. You should give them the chance to say that they don't know (Schuman & Presser 1981: 113–160).

Because there are so many floaters in the typical survey sample, the decision to include an explicit "Don't know" option for a question is important, especially with surveys of less educated populations. "Don't know" responses are chosen more often by those with less education (Schuman & Presser 1981: 113–146). Unfortunately, the inclusion of an explicit "Don't know" response choice also allows some people who *do* have a preference to take the easy way out and choose "Don't know."

Exhibit 7.1 Filter Questions and Skip Patterns


Filter question
14. Are you currently employed? YES NO
Skip pattern
If you answered NO to Question 14, please skip to Question 16.
If you answered YES to Question 14, please answer Question 15.
Contingent question
15. How satisfied are you with your current job?
Very satisfied
Somewhat satisfied
Not very satisfied
Not at all satisfied
16. How satisfied are you with your life in general?
Very satisfied
Somewhat satisfied
Not very satisfied
Not at all satisfied

Fence-sitters, people who see themselves as being neutral, may skew the results if you force them to choose between opposites. In most cases, about 10% to 20% of respondents—those who do not have strong feelings on an issue—will choose an explicit middle, neutral alternative (Schuman & Presser 1981: 161–178). Adding an explicit neutral response option is appropriate when you want to find out who is a fence-sitter.

Fence-sitting and floating can be managed by including an explicit “no opinion” category after all the substantive responses. If neutral sentiment is a possibility, also include a neutral category in the middle of the substantive responses (such as “neither agree nor disagree”) (Dillman 2000: 58–60). Finally, adding an open-ended question in which respondents are asked to discuss their opinions (or reasons for having no opinion) can help by shedding some light on why some persons choose “Don’t know” in response to a particular question (Smith 1984).

Fence-sitters: Survey respondents who see themselves as being neutral on an issue and choose a middle (neutral) response that is offered.

Make Response Categories Exhaustive and Mutually Exclusive

Questions with fixed response choices must provide one and only one possible response for everyone who is asked the question. First, all of the possibilities should be offered (choices should be *exhaustive*). In one survey of employees who were quitting their jobs at a telecommunications company, respondents were given these choices for "Why are you leaving [the company]?": (a) poor pay, (b) poor working environment, (c) poor benefits, or (d) poor relations with my boss. Clearly, there may be other reasons (e.g., family or health reasons, geographical preferences) to leave an employer. The response categories were not exhaustive. Or when asking college students their class (senior, junior, etc.), you should probably consider having an "other" category for nontraditional matriculants who may be on an unusual track.

Second, response choices shouldn't overlap—they should be mutually exclusive so that picking one rules out picking another. If I say, for instance, that I'm 25 years old, I cannot also be 50 years old, but I may claim to be both "young" and "mature." Those two choices aren't mutually exclusive, so they shouldn't be used as response categories for a question about age.

There are two exceptions to these principles: Filter questions may tell some respondents to skip over a question (the response choices do not have to be exhaustive), and respondents may be asked to "check all that apply" (the response choices are not mutually exclusive). Even these exceptions should be kept to a minimum. Respondents to a self-administered questionnaire should not have to do a lot of skipping around, or else they may lose interest in completing carefully all the applicable questions. And, some survey respondents react to a "check all that apply" request by just checking enough responses so that they feel they have "done enough" for that question and then ignoring the rest of the choices (Dillman 2000: 63).

Interactive Exercises

Survey Research

How Should Questionnaires Be Designed?

Survey questions are asked as part of a **questionnaire**—or **interview schedule**, in interview-based studies; they are not isolated from other questions. The context the questionnaire creates as a whole has a major impact on how individual questions are interpreted and answered. Therefore, survey researchers must carefully design the questionnaire itself, not just each individual question. Several steps, explained in the following sections, will help you design a good questionnaire.

Build on Existing Instruments

If another researcher has already designed a set of questions to measure a key concept and previous surveys indicate that this measure is reliable and valid, then by all means use that instrument. Resources such as the *Handbook of Research Design and Social Measurement, Sixth Edition* (Miller & Salkind 2002) can give you many ideas about existing questionnaires; your literature review at the start of a research project should be an even better source.

But there is a trade-off here. Questions used previously may not concern quite the right concept or may not be appropriate in some ways for your population. A good rule of thumb is to use a previously designed instrument if it measures the concept of concern to you and it seems appropriate for your survey population.

Questionnaire: A survey instrument containing the questions in a self-administered survey.

Interview schedule: A survey instrument containing the questions asked by the interviewer in an in-person or phone survey.

Refine and Test Questions

The only good question is a pretested question. Before you rely on a question in your research, you need evidence that your respondents will understand what it means. So try it out on a few people (Dillman 2000: 140–147).

One important form of pretesting is discussing the questionnaire with colleagues. You can also review prior research in which your key questions or indexes have been used. Another increasingly popular form of pretesting comes from guided discussions among potential respondents. Such *focus groups* let you check for consistent understanding of terms and identify the range of events or experiences about which people will be asked to report (Fowler 1995). (See Chapter 9 for more about this technique.)

Professional survey researchers have also developed a technique for evaluating questions called the **cognitive interview** (Fowler 1995). Although the specifics vary, the basic approach is to ask people to "think aloud" as they answer questions. The researcher asks a test question and then probes with follow-up questions to learn how the question was understood and whether its meaning varied for different respondents. This method can identify many potential problems.

Conducting a pilot study is the final stage of questionnaire preparation. For the Muslim America study, students placed 550 telephone calls and in the process learned (1) the extent of fear that many respondents felt about such a poll; (2) that females were, for cultural reasons, less likely to respond in surveys of the Muslim population; and (3) that some of their questions were worded ambiguously.

To do a pilot study, draw a small sample of individuals from the population you are studying or one very similar to it (it is best to draw a sample of at least 100 respondents) and carry out the survey procedures with them. You may include in the pretest version of a written questionnaire some space for individuals to add comments on each key question or, with in-person interviews, audiotape the test interviews for later review. Review the distribution of responses to each question, and revise any that respondents do not seem to understand.

A survey researcher also can try to understand what respondents mean by their

responses after the fact—that is, by including additional questions in the survey itself. Adding such **interpretive questions** after key survey questions is always a good idea, but it is of utmost importance when the questions in a survey have not been thoroughly pretested (Labaw 1980).

Cognitive interview: A technique for evaluating questions in which researchers ask people test questions, and then probe with follow-up questions to learn how they understood the question and what their answers mean.

Interpretive questions: Questions included in a questionnaire or interview schedule to help explain answers to other important questions.

Maintain Consistent Focus

A survey (with the exception of an omnibus survey) should be guided by a clear conception of the research problem under investigation and the population to be sampled. Remember to have measures of all of the independent and dependent variables you plan to use. Of course, not even the best researcher can anticipate the relevance of every question. Researchers tend to try to avoid “missing something” by erring on the side of extraneous questions (Labaw 1980: 40).

At the same time, long lists of redundant or unimportant questions dismay respondents, so respect their time and make sure that each question counts. Surveys too often include too many irrelevant questions.

Order the Questions

The sequence of questions on a survey matters. As a first step, the individual questions should be sorted into broad thematic categories, which then become separate sections in the questionnaire. Both the sections and the questions within the sections must then be organized in a logical order that would make sense in a conversation.

The first question deserves special attention, particularly if the questionnaire is to be self-administered. This question signals to the respondent what the survey is about, whether it will be interesting, and how easy it will be to complete ("Overall, would you say your physical health right now is excellent, good, fair, or poor?"). The first question should be connected to the primary purpose of the survey, it should be interesting, it should be easy, and it should apply to everyone in the sample (Dillman 2000: 92–94). Don't try to jump right into sensitive issues ("In general, how well do you think your marriage is working?"); respondents have to "warm up" before they will be ready for such questions. As a standard practice, for instance, most researchers ask any questions about income or finances near the end of a survey because many people are cautious about discussing such matters.

Question order can lead to **context effects** when one or more questions influence how subsequent questions are interpreted (Schober 1999: 89–98). The potential for context effects is greatest when two or more questions concern the same issue or closely related issues. For example, if an early question asks respondents to state for whom they plan to vote in an election, they may hesitate in later questions to support views that are clearly not those of that candidate. In general, people try to appear consistent (even if they are not); be sensitive to this and realize that earlier questions may "commit" respondents to answers on later questions.

Context effects: In survey research, refers to the influence that earlier questions may have on how subsequent questions are answered.

Careers and Research

Floyd J. (“Jack”) Fowler Jr., PhD, Founder and Director of the Center for Survey Research

Source: Floyd J. (“Jack”) Fowler Jr.

Jack Fowler “wrote the book” on survey research—two books, actually, with SAGE Publications: *Improving Survey Questions* (1995) and *Survey Research Methods (1988)*. This career focus crept up on Fowler while he was in school. As an undergraduate major in English at Wellesley College, Fowler found himself fascinated by social science and went on to earn his PhD in social psychology at the University of Michigan. In graduate school, he got hooked on survey research.

Fowler was asked to serve as a research assistant in a series of studies designed to identify the sources of error in the National Health Interview Survey, a major source of health data in the United States. This was an opportunity to relate research methodology to real-life problems and improve the way the world works, and Fowler seized it. He went on to found the Center for Survey Research at the University of Massachusetts Boston and to serve as its director for more than two decades. Fowler describes his professional life as “essentially a series of projects” that have made a difference by helping address important problems in areas ranging from health care, crime, and housing to medical decision making and views of local government.

His advice for students interested in a similar career:

> Methods, methods, methods. Make sure you are firmly grounded in the methods of collecting and analyzing data. The research priorities will change, society and the nature of the problems change so fast. However, if you know how to collect and analyze data, you will always be relevant. . . . To enjoy work most days and to feel confident it is making a positive difference is about as good as one can ask for.

Make the Questionnaire Attractive

An attractive questionnaire—neat, clear, clean, and spacious—is more likely to be completed and less likely to confuse either the respondent or, in an interview, the interviewer.

An attractive questionnaire does not look cramped; plenty of white space—more between questions than within question components—makes the questionnaire appear easy to complete. Response choices are listed vertically and are distinguished clearly and consistently, perhaps by formatting them in all capital letters and keeping them in the middle of the page. Skip patterns are indicated with arrows or other graphics. Some distinctive type of formatting should be used to identify instructions. Printing a multipage questionnaire in booklet form usually results in the most attractive and simple-to-use questionnaire (Dillman 2000: 80–86).

Exhibit 7.2 contains portions of a telephone interview questionnaire that illustrates these features, making it easy for the interviewer to use.

Exhibit 7.2 Sample Interview Guide

Hi, my name is _____. I am calling on behalf of (I am a student at) Hamilton College in New York. We are conducting a national opinion poll of high school students.

SCREENER: Is there a sophomore, junior, or senior in high school in your household with whom I may speak?

1. Yes 2. No/not sure/refuse **(End)**

(If student not on phone, ask:) Could he or she come to the phone?

(When student is on the phone) Hi, my name is _____. I am calling on behalf of (I am a student at) Hamilton College in New York. We are conducting a national opinion poll of high school students about gun control. Your answers will be completely anonymous. Would you be willing to participate in the poll?

1. Yes 2. No/not sure/refuse **(End)**

1. (SKOLYR) What year are you in school?

1. Sophomore
2. Junior
3. Senior
4. Not sure/refuse **(do not read) (End)**

Now some questions about your school:

2. (SKOL) Is it a public, Catholic, or private school?

1. Public 2. Catholic 3. Private 4. Not sure **(do not read)**

Source: Gilbert, Dennis (with Zogby International). 2000. Hamilton College youth and guns survey. Unpublished research report.

What Are the Alternatives for Administering Surveys?

Surveys can be administered in at least five different ways. They can be *mailed* or *group-administered* or conducted *by telephone*, *in person*, or *electronically.* (Exhibit 7.3 summarizes the typical features of each.) Each approach differs from the others in one or more important features:

Exhibit 7.3 Typical Features of the Five Survey Designs

Design	Manner of Administration	Setting	Questionnaire Structure	Cost
Mailed survey	Self	Individual	Structured	Low
Group survey	Self	Group	Structured	Very low
Phone survey	Professional	Individual	Structured	Moderate
In-person interview	Professional	Individual or unstructured	Mostly Structured	High
Electronic survey	Self	Individual	Structured	Very low

- *Manner of administration*—The respondents themselves complete mailed, group, and electronic surveys. During phone and in-person interviews, however, the researcher or a staff person asks the questions and records the respondent's answers.
- *Questionnaire structure*—Most mailed, group, phone, and electronic surveys are highly structured, fixing in advance the content and order of questions and response choices. In-person interviews may be highly structured, but they also may include many questions without fixed response choices.
- *Setting*—Mailed, electronic, and phone interviews are usually intended for only one respondent. The same is usually true of in-person interviews, although sometimes researchers interview several family members at once. However, some surveys are distributed simultaneously to a group of respondents, who complete the survey while the researcher (or assistant) waits.
- *Cost*—As mentioned earlier, in-person interviews are clearly the most expensive type of survey. Phone interviews are much less expensive, and surveying by mail is cheaper yet. Electronic surveys are now the least

expensive method, because there are no interviewer costs; no mailing costs; and, for many designs, almost no costs for data entry. (Of course, extra staff time and expertise are required to prepare an electronic questionnaire.)

Because of their different features, the five administrative options vary in the types of error to which they are most prone and the situations in which they are most appropriate. The rest of this section focuses on each format’s unique advantages and disadvantages.

Mailed, Self-Administered Surveys

A **mailed (self-administered) survey** is conducted by mailing a questionnaire to respondents, who then take the survey by themselves. The central problem for a mailed survey is maximizing the response rate. Even an attractive questionnaire with clear questions will probably be returned by no more than 30% of a sample unless extra steps are taken. A response rate of 30%, of course, is a disaster, destroying any hope of a representative sample. That's because people who *do* respond are often systematically different from people who *don't* respond—women respond more often, for instance, to most surveys; people with very strong opinions respond more than those who are indifferent; very wealthy and very poor people, for different reasons, are less likely to respond.

Fortunately, the conscientious use of systematic techniques can push the response rate to 70% or higher for most mailed surveys (Dillman 2000: 27), which is acceptable. Sending follow-up mailings to nonrespondents is the single most important technique for obtaining an adequate response rate. The follow-up mailings explicitly encourage initial nonrespondents to return a completed questionnaire; implicitly, they convey the importance of the effort. Dillman (pp. 155–158, 177–188) has demonstrated the effectiveness of a standard procedure for the mailing process: a preliminary introductory letter, a well-packaged survey mailing with a personalized **cover letter**, a reminder postcard 2 weeks after the initial mailing, and then new cover letters and replacement questionnaires 2 to 4 weeks and 6 to 8 weeks after that mailing.

Video Link

Watch a video about survey design in a specific context.

The cover letter, actually, is critical to the success of a mailed survey. This statement to respondents sets the tone for the entire questionnaire. The cover letter or introductory statement must establish the credibility of the research and the researcher, it must be personalized (including a personal salutation and an original signature), it should be interesting to read, and it must explain issues about voluntary participation and maintaining subject confidentiality (Dillman 1978: 165–172). A carefully prepared cover letter should increase the response

rate and result in more honest and complete answers to the survey questions; a poorly prepared cover letter can have the reverse effects. Exhibit 7.4 is an example of a cover letter for a questionnaire.

Exhibit 7.4 Sample Questionnaire Cover Letter

University of Massachusetts Boston
Department of Sociology
May 24, 2014

Jane Doe
AIDS Coordinator
Shattuck Shelter

Dear Jane:

AIDS is an increasing concern for homeless people and for homeless shelters. The enclosed survey is about the AIDS problem and related issues confronting shelters. It is sponsored by the Life Lines AIDS Prevention Project for the Homeless—a program of the Massachusetts Department of Public Health.

As an AIDS coordinator/shelter director, you have learned about homeless persons' problems and about implementing programs in response to those problems. The Life Lines Project needs to learn from your experience. Your answers to the questions in the enclosed survey will improve substantially the base of information for improving AIDS prevention programs.

Questions in the survey focus on AIDS prevention activities and on related aspects of shelter operations. It should take about 30 minutes to answer all the questions.

Every shelter AIDS coordinator (or shelter director) in Massachusetts is being asked to complete the survey. And every response is vital to the success of the survey: The survey report must represent the full range of experiences.

You may be assured of complete confidentiality. No one outside of the university will have access to the questionnaire you return. (The ID number on the survey will permit us to check with nonrespondents to see if they need a replacement survey or other information.) All information presented in the report to Life Lines will be in aggregate form, with the exception of a list of the number, gender, and family status of each shelter's guests.

Please mail the survey back to us by Monday, June 9, and feel free to call if you have any questions.

Thank you for your assistance.

Yours sincerely,

Russell K. Schutt
Russell K. Schutt, PhD
Project Director

Stephanie Howard
Stephanie Howard
Project Assistant

Audio Link

Listen to new ways researchers are getting survey data.

Other steps that help to maximize the response rate include clear and

understandable questions, not many open-ended questions, a credible research sponsor, a token incentive (such as a $1 coupon), and presurvey advertising (Fowler 1988: 99–106; Mangione 1995: 79–82).

Mailed (self-administered) survey: A survey involving a mailed questionnaire to be completed by the respondent.

Cover letter: The letter sent with a mailed questionnaire that explains the survey's purpose and auspices and encourages the respondent to participate.

Group-Administered Surveys

A **group-administered survey** is completed by individual respondents assembled in a group. The response rate is usually high because most group members will participate. Unfortunately, this method is seldom feasible because it requires a captive audience. With the exception of students, employees, members of the armed forces, and some institutionalized populations, most people cannot be sampled in such a setting.

Whoever is responsible for administering the survey to the group must be careful to minimize comments that might bias answers or that could vary between different groups in the same survey (Dillman 2000: 253–256). A standard introductory statement should be read to the group that expresses appreciation for their participation, describes the steps of the survey, and emphasizes (in classroom surveys) that the survey is not the same as a test. A cover letter like that used in mailed surveys also should be distributed with the questionnaires. To emphasize confidentiality, respondents should be given envelopes in which to seal their questionnaires after they are completed.

Another issue of special concern with group-administered surveys is the possibility that respondents will feel coerced to participate and, therefore, will be less likely to answer questions honestly. Also, because administering group surveys requires approval of the authorities—and this sponsorship is made quite obvious because the survey is conducted on the organization's premises—respondents may infer that the researcher is in league with the sponsor. No complete solution to this problem exists, but it helps to make an introductory statement emphasizing the researcher's independence and giving participants a chance to ask questions about the survey. The sponsor should keep a low profile and allow the researcher both control over the data and autonomy in report writing.

Journal Link

Read about a study that uses telephone surveys.

Group-administered survey: A survey that is completed by individual respondents who are

assembled in a group.

Telephone Surveys

In a **phone survey**, interviewers question respondents over the phone and then record respondents' answers. Phone interviewing is traditionally a very popular method of conducting surveys in the United States because almost all families have phones. But two problems often threaten the validity of a phone survey: not reaching the proper sampling units (or *coverage error*) and not getting enough successfully completed responses to make the results generalizable.

Reaching Sampling Units

The first big problem lies in the difficulty of actually contacting the sample units (typically households). Most telephone surveys use random digit dialing (RDD) at some point in the sampling process (Lavrakas 1987) to contact a random sample of households. A machine calls random phone numbers within the designated exchanges, whether or not the numbers are published. RDD is a good way to "capture" unlisted numbers, whose owners are systematically different (often they are wealthier than the general population). When the machine reaches an inappropriate household (such as a business, in a survey of individuals), the phone number is simply replaced with another.

But the tremendous recent (since 2000) popularity of cellular, or mobile, telephones (and now smartphones) has made accurate coverage of random samples almost impossible, for several reasons (Tavernise 2011: A13; Tourangeau 2004: 781–792): (1) Cell phones are typically not listed in telephone directories, so they can't be included in prepared calling lists; (2) close to 27% of the U.S. population now has only a cell phone (no landline) and therefore can't be reached by either RDD or many directories; and (3) for 18- to 30-year-olds, some 44% have cell phones only. Cell-phone-only households are also more common among non-English speakers and among poor people.

Researcher Interview Link

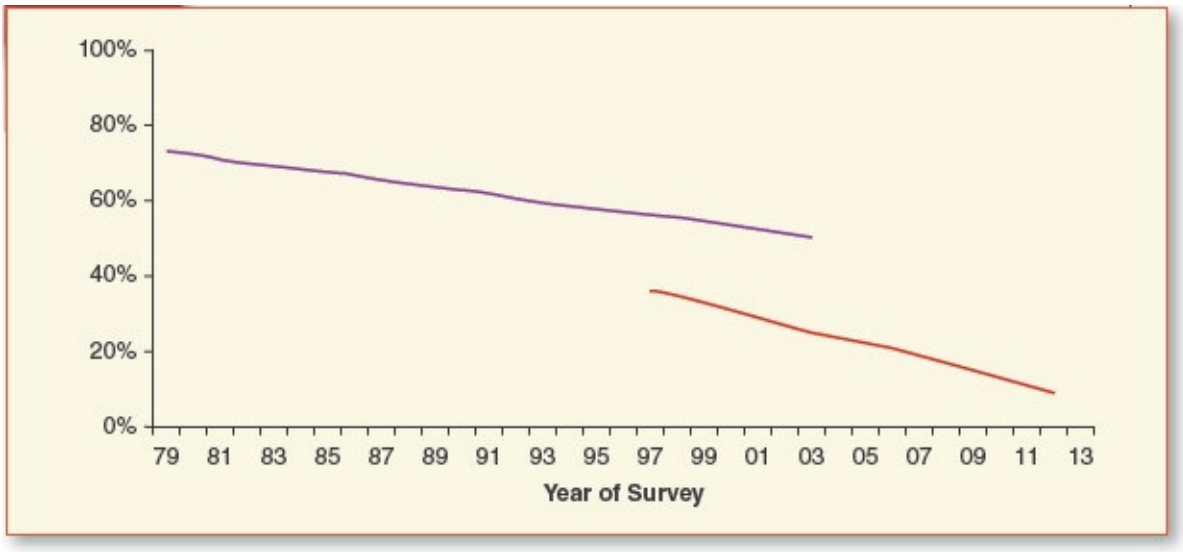
Watch a researcher describe using telephone surveys in order to receive valid estimates of sexually transmitted infection prevalence rates.

The net effect, then, of widespread cell phone usage is to underrepresent young, poor, and non-English-speaking people in particular from inclusion in most large telephone surveys, obviously damaging the results.

Even if an appropriate (for sampling) number is dialed, responses may not be completed. Because people often are not home, multiple callbacks will be needed for many sample members. With large numbers of single-person households, dual-earner families, and out-of-home activities, survey research organizations have had to increase the usual number of phone contact attempts from just 4 to 8 tries to 20—a lot of attempts just to reach one person. Those with more money and education are more likely to be away from home; such persons are more likely to vote Republican, so the results of political polls can be seriously biased if few callback attempts are made (Kohut 1988). This problem has been compounded in recent years by social changes that are lowering the response rate in phone surveys (Tourangeau 2004: 781–783) (see Exhibit 7.5). The Pew Research Center reports a decline in the response rate based on all those sampled from 36% in 1997 to only 9% in 2012 (Kohut et al. 2012).

Response rates can also be much lower in harder-to-reach populations (Exhibit 7.5). In a recent phone survey of low-income women in a public health program (Schutt & Fawcett 2005), the University of Massachusetts Center for Survey Research (CSR) achieved a 55.1% response rate from all eligible sampled clients after a protocol that included as many as 30 contact attempts, although the response rate rose to 72.9% when it was calculated as a percentage of clients who CSR was able to locate (Roman 2005: 7). Caller ID and call waiting allow potential respondents to avoid answering calls from strangers, including researchers. The growth of telemarketing has accustomed individuals nowadays to refuse calls from unknown individuals and organizations or to use their answering machines to screen calls (Dillman 2000: 8, 28).

Exhibit 7.5 Phone Survey Response Rates by Year, 1979–2012



Source: Adapted from R. Curtin, S. Presser, and E. Singer. 2005. Changes in telephone survey nonresponse over the past quarter century. Public Opinion Quarterly 69:87–98. Copyright © 2005, Oxford University Press, on behalf of the American Association for Public Opinion Research. Reprinted with permission.

After all, respondents don't really know who is calling and may have good reason to be suspicious. In the Muslim America study, many people were afraid to talk with the researchers or were actively hostile. Finally, a huge number of cell phone users are children, and therefore legally unavailable for surveys, so calls made to them are all wasted efforts for researchers.

Taken together, this huge range of problems means that careful training and direction of interviewers is essential in phone surveys. The instructions shown in Exhibit 7.6 were developed to clarify procedures for asking and coding a series of questions in the phone interviews conducted for the youth and guns survey.

Phone surveying is the method of choice for relatively short surveys of the general population. Response rates in phone surveys traditionally have tended to be very high—often above 80%—because few individuals would hang up on a polite caller or refuse to stop answering questions (at least within the first 30 minutes or so). But the problems we have noted, especially those connected with cell phone usage, makes this method of surveying populations increasingly difficult. The long-term decline in response rates to household surveys is such a

problem for survey researchers that they have devoted entire issues of major journals to it (Singer 2006: 637–645). Traditionally, a high response rate because it preserves the sample selected has been considered preferable. But given the difficulty nowadays of getting responses for some people, it may be that high response rates may themselves—oddly enough—introduce bias: If someone is so difficult to persuade, they may not be a typical person. And in certain cases, it's not clear that low response rates actually bias the sample. Sophisticated professionals differ over these issues. But usually, a high response is better overall.

Research|Social Impact Link

Read about how campus surveys may help to understand important safety issues.

Exhibit 7.6 Sample Interviewer Instructions

22. (CONSTIT) To your knowledge, does the U.S. Constitution guarantee citizens the right to own firearms?

1. Yes 2. No (**skip to 24**) 3. Not sure (**do not read**)

23. (CONLAW) Do you believe that laws regulating the sale and use of handguns violate the constitutional rights of gun owners?

1. Yes 2. No 3. Not sure (**do not read**)

24. (PETITION) In some localities, high school students have joined campaigns to change the gun laws, and sometimes they have been successful. Earlier you said that you thought that the current gun control laws were (**if Q11 = 1, insert "not strict enough"; if Q11 = 2, insert "too strict"**). Suppose a friend who thinks like you do about this asked you to sign a petition calling for (**if Q11 = 1, insert "stronger gun control laws"; if Q11 = 2, insert "less restrictive gun control laws"**). On a scale from 1 to 5, with 1 being very unlikely and 5 being very likely, how likely is it that you would sign the petition?

1. (Very unlikely)
2.
3.
4.
5. (Very likely)
6. Not sure (do not read)

Source: Gilbert, Dennis (with Zogby International). 2000. Hamilton

College youth and guns survey. Unpublished research report.

An interesting variant of telephone surveys that you may have encountered is the IVR survey. Computerized **interactive voice response (IVR)** survey technology allows great control over interviewer-respondent interaction. In an IVR survey, respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software. These surveys can also record verbal responses to open-ended questions for later transcription. Although they present some difficulties when many answer choices must be used or skip patterns must be followed, IVR surveys have been used successfully with short questionnaires and when respondents are highly motivated to participate (Dillman 2000: 402–411). When these conditions are not met, potential respondents may be put off by the impersonality of this computer-driven approach.

Phone survey: A survey in which interviewers question respondents over the phone and record their answers.

Interactive voice response (IVR): A survey in which respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software.

In-Person Interviews

What is unique to the **in-person interview**, compared with the other survey designs, is the face-to-face social interaction between interviewer and respondent. If money is no object, in-person interviewing is often the best survey design.

In-person interviewing has several advantages: Response rates are higher than with any other survey design; questionnaires can be much longer than with mailed or phone surveys; the questionnaire can be complex, with both open-ended and closed-ended questions and frequent branching patterns; the interviewer can control the order in which questions are read and answered; the physical and social circumstances of the interview can be monitored; and respondents' interpretations of questions can be probed and clarified. The interviewer, therefore, is well placed to gain a full understanding of what the respondent really wants to say.

Researcher Interview Link

Watch a researcher describe using interviews to understand how women experience reintegration into a community after incarceration.

However, researchers must be alert to some special hazards resulting from the presence of an interviewer. Ideally, every respondent should have the same interview experience—that is, each respondent should be asked the same questions in the same way by the same type of person, who reacts similarly to the answers. Suppose one interviewer is smiling and pleasant while another is gruff and rude; the two interviewers will likely elicit very different results in their surveys, if only in the length of responses. Careful training and supervision are essential (Groves 1989: 404–406).

Computers can be used to increase control of the in-person interview. In a **computer-assisted personal interview (CAPI)** project, interviewers carry a laptop computer that is programmed to display the interview questions and to process the responses that the interviewer types in, as well as to check that these responses fall within allowed ranges (Tourangeau 2004: 790–791). Interviewers seem to like CAPI, and the data obtained are comparable in quality to data

obtained in a noncomputerized interview (Shepherd et al. 1996). A CAPI approach also makes it easier for the researcher to develop skip patterns and experiment with different types of questions for different respondents without increasing the risk of interviewer mistakes (Couper et al. 1998).

The presence of an interviewer may make it more difficult for respondents to give honest answers to questions about socially undesirable behaviors such as drug use, sexual activity, and not voting (Schaeffer & Presser 2003: 75). CAPI is valued for this reason because respondents can enter their answers directly in the laptop without the interviewer knowing what their response is. Alternatively, interviewers can simply hand respondents a separate self-administered questionnaire containing the more sensitive questions. After answering those questions, the respondent seals the separate questionnaire in an envelope so that the interviewer does not know the answers. When this approach was used for the GSS questions about sexual activity, about 21% of men and 13% of women who were married or had been married admitted to having cheated on a spouse ("Survey on Adultery" 1993: A20).

Maximizing Response to Interviews

Several factors affect the response rate in interview studies. Contact rates tend to be lower in central cities, in part because of difficulties in finding people at home and gaining access to high-rise apartments, and, in part, because of interviewer reluctance to visit some areas at night, when people are more likely to be home (Fowler 1988: 45–60). Households with young children or elderly adults tend to be easier to contact, whereas single-person households are more difficult to reach (Groves & Couper 1998: 119–154).

Encyclopedia Link

Read an overview of how response rates have changed over time.

Refusal rates vary with some respondent characteristics. People with less education participate somewhat less in surveys of political issues (perhaps because they are less aware of current political issues). Less education is also associated with higher rates of "Don't know" responses (Groves 1989). Conversely, wealthy people often refuse to be surveyed about their income or

buying habits, perhaps to avoid being plagued by sales calls. Such problems can be lessened with an advance letter introducing the survey project and by multiple contact attempts throughout the day and evening, but they cannot be entirely avoided (Fowler 1988: 52–53).

In-person interview: A survey in which an interviewer questions respondents face-to-face and records their answers.

Computer-assisted personal interview (CAPI): A personal interview in which the laptop computer is used to display interview questions and to process responses that the interviewer types in, as well as to check that these responses fall within allowed ranges.

Research That Matters

When older people become more physically impaired in daily life, they suffer more psychological distress; being married, however, is associated with less psychological distress. But can marriage actually reduce some of the adverse effects of aging? Alex Bierman at the University of Calgary sought to answer this question with data collected in the Aging, Status, and Sense of Control survey (ASOC) by the Survey Research Laboratory at the University of Illinois at Urbana-Champaign. ASOC used a longitudinal panel design that attempted to survey the same respondents in 1995, 1998, and 2001. There were 966 respondents aged 60 and older in the first wave and 907 who participated in all three waves.

The ASOC measured psychological distress with responses to seven questions. The questions ask about the number of times in the previous week the respondent had trouble getting to sleep, felt that everything was an effort, felt they couldn't get going, and so on. "Functional limitations" (impairment) was measured with respondents' ratings of difficulty with such tasks as climbing stairs, kneeling, and shopping. Bierman's analysis indicated that married respondents tended to feel less psychological distress in relation to functional limitations they experienced than did unmarried respondents—and this "protective effect" of marriage was stronger for men than for women.

Source: Adapted from Bierman, Alex. 2012. Functional limitations and psychological distress: Marital status as moderator. *Society and Mental Health* 2(1): 35–52.

Electronic Surveys

The widespread use of personal computers and the growth of the Internet have created new possibilities for survey research. **Electronic surveys** can be prepared in two ways (Dillman 2000: 352–354). **E-mail surveys** can be sent as messages to respondent e-mail addresses. Respondents then mark their answers in the message and send them back to the researcher. This approach is easy for researchers to develop and for respondents to use. However, this approach is cumbersome for surveys that are more than four or five pages long. **Web surveys** are stored on a server that the researcher controls; respondents are then asked to visit the website (often by just clicking an e-mailed link) and respond to the questionnaire by checking answers. Web surveys require more programming by the researcher, but a well-designed web survey can tailor its questions to a given respondent and thus seem shorter, more interesting, and more attractive.

Web surveys have recently become increasingly useful and popular for two reasons: growth in the percentage of people using the Internet, and technological advances that make web design relatively easy. Many specific populations have very high rates of Internet use, so a web survey can be a good option for groups such as professionals, middle-class communities, members of organizations, and of course, college students. Because of the Internet's global reach, web surveys also make it possible to conduct large, international surveys. However, coverage remains a major problem with many populations (Tourangeau, Conrad, & Couper 2012). About one quarter of U.S. households are not connected to the Internet (File 2013b), so it is not yet possible to survey directly a representative sample of the U.S. population on the web—and given a plateau in the rate of Internet connections, this coverage problem may persist for the near future (Couper & Miller 2008: 832). Rates of Internet usage are much lower in other parts of the world, with a worldwide average of 34.3% and rates as low as 15.6% in Africa and 27.5% averaged across all of Asia (Internet World Statistics 2012). Households without Internet access also tend to be older, poorer, and less educated than do those that are connected, so web surveys of the general population can result in seriously biased estimates (File 2013b; Pew Research Center 2013). Coverage problems can be compounded in web surveys because of much lower rates of survey completion: It is just too easy to stop working on a web survey—much easier than it is to break off interaction with an interviewer (Tourangeau et al. 2012).

Web surveys can help in obtaining a large sample, getting rapid turnaround, collecting sensitive information that might be embarrassing to acknowledge in person, using an email list of the population, and employing interactive and multimedia features will enhance interest in the survey (Sue & Ritter 2012: 10–11). Jennie Connor, Andrew Gray, and Kypros Kypri (2010) achieved an impressive 63% response rate with a web survey about substance use that began with an initial e-mail invitation to a representative sample of undergraduate students at six New Zealand campuses.

There are several different approaches to engaging people in web surveys. Many web surveys begin with an e-mail message to potential respondents that contains a direct "hotlink" to the survey website (Gaiser & Schreiner 2009: 70). Such e-mail invitations should include a catchy phrase in the subject line, as well as attractive and clear text in the message itself (Sue & Ritter 2012: 110–114). This approach works well when a defined population with known e-mail addresses is to be surveyed. The researcher can then send e-mail invitations to a representative sample without difficulty (Dillman 2000: 378; Sue & Ritter 2012: 103–104). Connor and colleagues (2010: 488) used this approach in their survey of New Zealand undergraduates.

However, lists of unique e-mail addresses for the members of defined populations generally do not exist outside of organizational settings. Many people have more than one e-mail address, and often there is no apparent link between an e-mail address and the name or location of the person to whom it is assigned. As a result, there is no available method for drawing a random sample of e-mail addresses for people from any general population, even if the focus is only on those with Internet access (Dillman 2007: 449).

Web surveys that use volunteer samples may instead be linked to a website that is used by the intended population and everyone who visits that site is invited to complete the survey. This was the approach used in the international web survey sponsored by the National Geographic Society in 2000 (Witte, Amoroso, & Howard 2000). However, although this approach can generate a very large number of respondents (50,000 persons completed Survey2000), the resulting sample will necessarily reflect the type of people who visit that website (middle class, young North Americans, in Survey 2000) and thus be a biased representation of the larger population (Couper 2000: 486–487; Dillman 2000: 355). Some control over the resulting sample can be maintained by requiring participants to meet certain inclusion criteria (Selm & Jankowski 2006: 440).

Journal Link

Read about a study that implements a web survey.

Coverage bias can even be a problem with web surveys aimed at populations with high levels of Internet use. If the topic of the survey prompts some people to be more likely to respond, the resulting sample can be very unrepresentative. William Wells and colleagues (2012: 461) identified this problem when comparing students responding to a web-based survey about gun violence with students—at the same university—who responded to the same survey when it was administered in classes. Students responding to the web survey were much more likely to support the right to carry concealed weapons on campus than were those who took the classroom survey, suggesting that the web's anonymity may have over-invited those with more extreme views.

Some web surveys are designed to reduce coverage bias by providing computers and Internet connections to those who do not have them. This design-based recruitment method begins by contacting people by phone and providing those who agree to participate with whatever equipment they lack. This approach considerably increases the cost of the survey, so it is normally used as part of creating the panel of respondents who agree to be contacted for multiple surveys over time. The start-up costs can then be spread across many surveys. Gfk Knowledge Networks is a company that received funding from the U.S. National Science Foundation to create such a web survey panel. CentER Data in the Netherlands also uses this panel approach (Couper & Miller 2008: 832–833). Another approach to reducing coverage bias in web surveys is to recruit a volunteer panel of Internet users and then weight the resulting sample to make it comparable to the general population in such demographics as gender, race, age, and education. This is the method adopted by many market research organizations (Couper & Miller 2008: 832–833); although response rates to volunteer samples are very low and the participants are often unlike the general population, it appears that weighting can reduce coverage bias by 30% to 60% (Tourangeau et al. 2012).

Sometimes a convenience sample will suffice for an exploratory survey about some topic. Audrey Freshman (2012: 41) used a web survey with a convenience sample to study symptoms of posttraumatic stress disorder (PTSD) among

victims of the Bernie Madoff financial scandal.

This convenience, nonprobability sample was solicited via direct link to the study placed in online Madoff survivor support groups and comment sections of newspapers and blogs dealing with the event. The study announcement encouraged victims to forward the link to other former investors who might be interested in responding to the survey, thereby creating a snowball effect (Freshman 2012: 41).

Although a majority of respondents met clinical criteria for a diagnosis of PTSD, there is no way to know if this sample represents the larger population of Madoff's victims.

In contrast to problems of coverage, web surveys have some unique advantages for increasing measurement validity (Selm & Jankowski 2006; Tourangeau et al. 2012). Questionnaires completed on the web can elicit more honest reports about socially undesirable behavior or experiences, including illicit behavior and victimization in the general population and failing course grades among college students, when compared with results with phone interviews (Kreuter, Presser, & Tourangeau 2008; Parks, Pardi, & Bradizza 2006). Jane Onoye and colleagues (2012) found that conducting a survey on the web increased self-reports of substance use compared with a paper-and-pencil survey. Web surveys are relatively easy to complete because respondents simply click on response boxes and the survey can be programmed to move respondents easily through sets of questions, not presenting questions that do not apply to the respondent, thus leading to higher rates of item completion (Kreuter et al. 2008). (See Exhibit 7.7.) Because answers are recorded directly in the researcher's database, data entry errors are almost eliminated and results can be reported quickly.

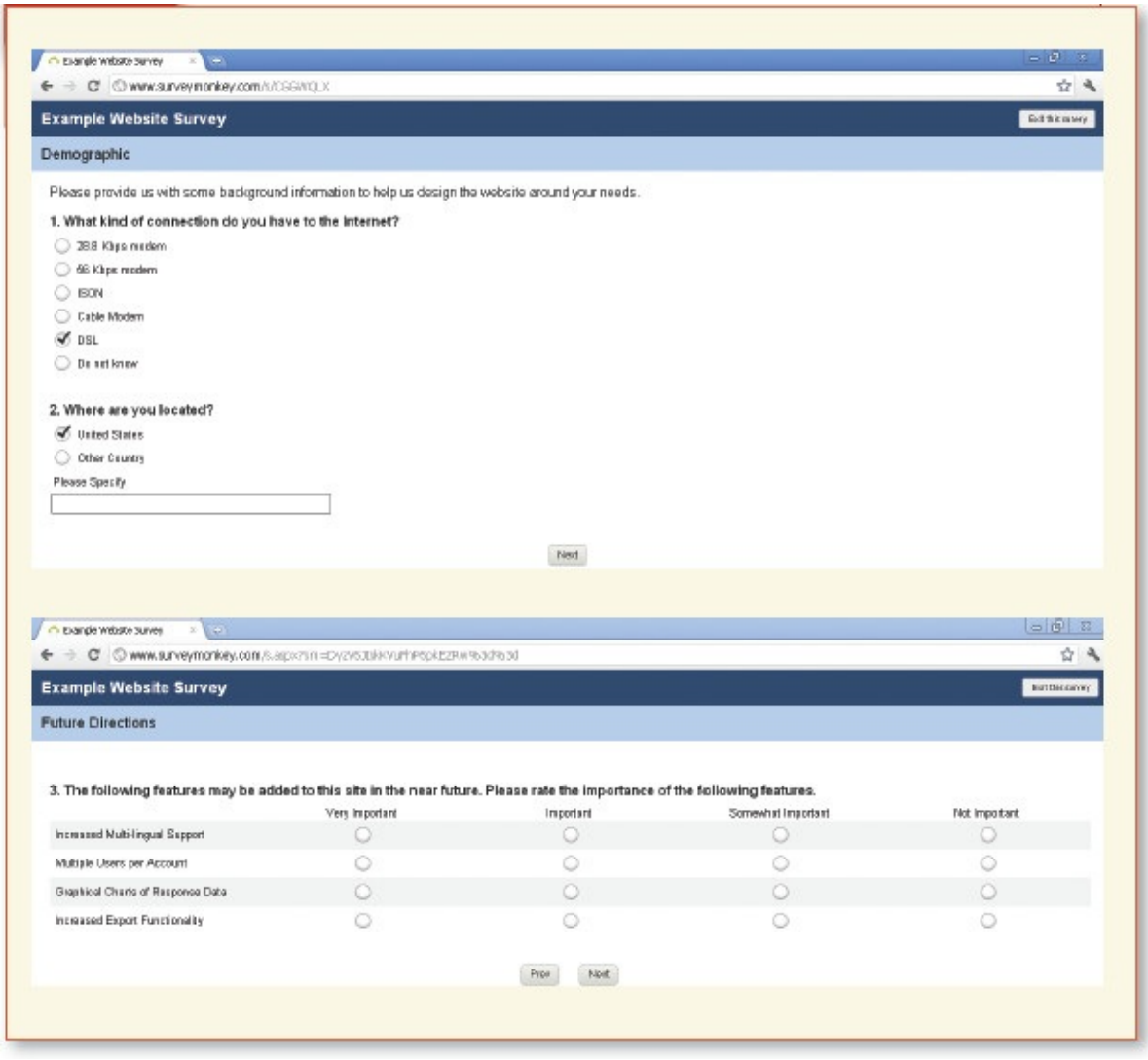
Despite some clear advantages of some types of web surveys, researchers who use this method must be aware of some important disadvantages. Coverage bias is the single biggest problem with web surveys of the general population and of segments of the population without a high level of Internet access, and none of the different web survey methods fully overcome this problem. Web surveys that take more than 15 minutes are too long for most respondents (de Leeuw 2008: 322). Surveys by phone continue to elicit higher rates of response (Kreuter et al. 2008). Some researchers have found that when people are sent a mailed survey that also provides a link to a web survey alternative, they overwhelmingly choose the paper survey (Couper 2000: 488).

Surveys are also now being conducted through social media such as Facebook, on smartphones, and via text messages (Sue & Ritter 2012: 119–122). Research continues into the ways that the design of web surveys can influence rates of initial response, the likelihood of completing the survey, and the validity of the responses (Couper, Traugott, & Lamias 2001; Kreuter et al. 2008; Porter & Whitcomb 2003; Tourangeau et al. 2012). At this point, there is reason enough to consider the option of a web survey for many investigations, but to proceed with caution and consider carefully their strengths and weaknesses when designing a web survey of any type and when analyzing findings from it.

Research|Social Impact Link

Read about how different survey methods can achieve different results.

Exhibit 7.7 Survey Monkey Web Survey Example



Electronic survey: A survey that is sent and answered by computer, either through e-mail or on the web.

E-mail survey: A survey that is sent and answered through e-mail.

Web survey: A survey that is accessed and responded to on the World Wide Web.

A Comparison of Survey Designs

Which survey design should you use for a study? Let's compare the four major survey designs: (1) mailed surveys, (2) phone surveys, (3) in-person surveys, and (4) electronic surveys. (Group-administered surveys are similar in most respects to mailed surveys except that they require the unusual circumstance of having access to the sample in a group setting.) Exhibit 7.8 summarizes these strong and weak points.

The most important difference among these four methods is their varying response rates. Because of the low response rates of *mailed surveys*, they are weakest from a sampling standpoint. However, researchers with limited time, money, and staff may still prefer a mailed survey. Mailed surveys can be useful in asking sensitive questions (e.g., questions about marital difficulties or financial situations), because respondents won't be embarrassed by answering in front of an interviewer.

Contracting with an established survey research organization for a *phone survey* is often the best alternative to a mailed survey. The persistent follow-up attempts that are necessary to secure an adequate response rate are much easier over the phone than in person, although you must be careful about the cell phone sampling and response problem. A phone survey limits the length and complexity of the questionnaire but offers the possibility of very carefully monitoring interviewers (Fowler 1988: 61–73).

Exhibit 7.8 Advantages and Disadvantages of Four Survey Designs

Characteristics of Design	Mail Survey	Phone Survey	In-Person Survey	Web Survey
Representative sample				
Opportunity for inclusion is known				
For completely listed populations	High	High	High	Medium
For incompletely listed populations	Medium	Medium	High	Low
Selection within sampling units is controlled (e.g., specific family members must respond)	Medium	High	High	Low
Respondents are likely to be located				
If samples are heterogeneous	Medium	High	High	Low
If samples are homogeneous and specialized	High	High	High	High
Questionnaire construction and question design				
Allowable length of questionnaire Ability to include	Medium	Medium	High	Medium
Complex questions	Medium	Low	High	High
Open questions	Low	High	High	Medium
Screening questions	Low	High	High	High
Tedious, boring questions	Low	High	High	Low
Ability to control question sequence	Low	High	High	High
Ability to ensure questionnaire completion	Medium	High	High	Low
Distortion of answers				
Odds of avoiding social desirability bias	High	Medium	Low	High
Odds of avoiding interviewer distortion	High	Medium	Low	High
Odds of avoiding contamination by others	Medium	High	Medium	Medium
Administrative goals				
Odds of meeting personnel requirements	High	High	Low	Medium
Odds of implementing quickly	Low	High	Low	High
Odds of keeping costs low	High	Medium	Low	High

Source: Adapted from Dillman, Don A. 1978. Mail and telephone surveys: The total design method. New York: Wiley. Copyright © 1978 by John Wiley & Sons, Inc. Reprinted with permission of John Wiley & Sons, Inc.

In-person surveys can be long and complex, and the interviewer can easily monitor the conditions (the room, noise and other distractions, etc.). Although interviewers may themselves distort results, either by changing the wording of questions or failing to record answers properly, this problem can be lessened by careful training and monitoring of interviewers and by tape-recording the

answers.

The advantages and disadvantages of *electronic surveys*, including *web surveys*, depend on the populations to be surveyed. Too many people do not have Internet connections for general use of Internet surveying. But when your entire sample has access and ability (e.g., college students, corporate employees), web-based surveys can be very effective.

So overall, in-person interviews are the strongest design and are generally preferable when sufficient resources and a trained interview staff are available; telephone surveys have many of the advantages of in-person interviews at much less cost, but coverage response rates are an increasing problem. Any decision about the best survey design for a particular study must consider the particular features and goals of the study.

Ethical Issues in Survey Research

Survey research designs usually pose fewer ethical dilemmas than do experimental or field research designs. Potential respondents to a survey can easily refuse to participate, and a cover letter or introductory statement that identifies the sponsors of and motivations for the survey gives them the information required to make this decision. Little is concealed from the respondents, and the methods of data collection are quite obvious. Only in group-administered survey designs might the respondents (such as students or employees) be, in effect, a captive audience, so they require special attention to ensure that participation is truly voluntary. (Those who do not wish to participate may be told they can just hand in a blank form.)

Sometimes, political or marketing surveys are used unscrupulously to sway opinion under the guise of asking for it. So-called push polls are sometimes employed in political campaigns to distort an opponent's image ("If you knew Congressman Jones was cheating on his wife, would you consider him fit for high office?"). Advertisers can use surveys that pretend to collect opinions or "register" a purchase for warranty purposes, but often they are really trying to collate information about where you live, your phone numbers, your buying habits, and the like.

Confidentiality is most often the primary focus of ethical concern in survey research. Many surveys include questions that might prove damaging to the subjects if their answers were disclosed. When a survey of employees asks, "Do you think management here, especially your boss, is doing a good job?" or when student course evaluations ask, "On a scale of 1 to 5, how fair would you say the professor is?" respondents may well hesitate; if the boss or professor saw the results, workers or students could be hurt.

To prevent any disclosure of such information, it is critical to preserve subject confidentiality. Only research personnel should have access to information that could be used to link respondents to their responses, and even that access should be limited to what is necessary for specific research purposes. Only numbers should be used to identify respondents on their questionnaires, and the researcher should keep the names that correspond to these numbers in a safe, private location, unavailable to staff and others who might come across them.

Trustworthy assistants under close supervision should carry out follow-up mailings or contact attempts that require linking the ID numbers with names and addresses. If an electronic survey is used, encryption technology should be used to make information that is provided over the Internet secure from unauthorized people. Usually confidentiality can be protected readily; the key is to be aware of the issue. Don't allow bosses to collect workers' surveys or professors to pick up course evaluations. Be aware of your respondents' concerns and be even a little more careful than you need to be.

Few surveys can provide true **anonymity**, where no identifying information is ever recorded to link respondents with their responses. The main problem with anonymous surveys is that they preclude follow-up attempts to contact nonrespondents and they prevent panel designs, which measure change through repeated surveys of the same individuals. In-person surveys rarely can be anonymous because an interviewer must, in almost all cases, know the name and address of the interviewee. However, phone surveys that are meant only to sample opinion at one point in time, as in political polls, can safely be completely anonymous. When no future follow-up is desired, group-administered surveys also can be anonymous. To provide anonymity in a mail survey, the researcher should omit identifying codes from the questionnaire but may include a self-addressed, stamped postcard, so the respondent can notify the researcher that the questionnaire has been returned without creating any linkage to the questionnaire itself (Mangione 1995: 69).

Anonymity: Provided by research in which no identifying information is recorded that could be used to link respondents to their responses.

Conclusion

Survey research is an exceptionally efficient and productive method for investigating a wide array of social research questions. In addition to the potential benefits for social science, considerations of time and expense frequently make a survey the preferred data-collection method. One or more of the five survey designs reviewed in this chapter can be applied to almost any research question. It is no wonder that surveys have become the most popular research method in sociology and that they frequently inform discussion and planning about important social and political questions. As use of the Internet increases, survey research should become even more efficient and popular.

The relative ease of conducting at least some types of survey research leads many people to imagine that no particular training or systematic procedures are required. Nothing could be further from the truth. But as a result of this widespread misconception, you will encounter a great many nearly worthless survey results. You must be prepared to examine carefully the procedures used in any survey before accepting its findings as credible. And if you decide to conduct a survey, you must be prepared to invest the time and effort required by proper procedures.

Key Terms

Anonymity 159
Cognitive interview 145
Computer-assisted personal interview (CAPI) 153
Context effects 146
Contingent question 143
Cover letter 149
Double-barreled question 141
Double negative 141
Electronic survey 154
E-mail survey 154
Fence-sitters 144
Filter question 143
Floaters 143
Group-administered survey 150
In-person interview 153
Interactive voice response (IVR) 153
Interpretive questions 145
Interview schedule 144
Mailed (self-administered) survey 148
Omnibus survey 140
Phone survey 150
Questionnaire 144
Skip pattern 143
Survey research 140
Web survey 154

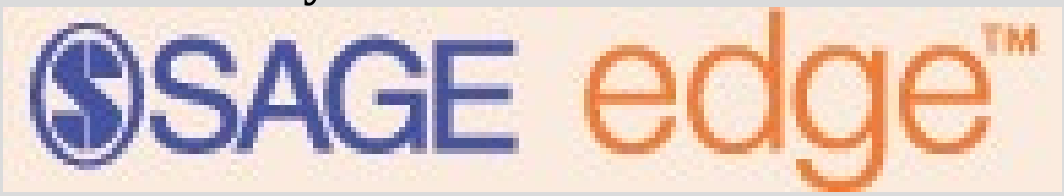
Highlights

- Surveys are the most popular form of social research because of their versatility, efficiency, and generalizability. Many survey data sets, such as the General Social Survey, are available for social scientists to use in teaching and research.
- Omnibus surveys cover a range of topics of interest and generate data useful to multiple sponsors.
- Questions must be worded carefully to avoid confusing respondents, encouraging less-than-honest responses, or triggering biases. Inclusion of "Don't know" choices and neutral responses may help, but the presence of such options also affects the distribution of answers. Open-ended questions can be used to determine the meaning that respondents attach to their answers. Answers to any survey questions may be affected by the questions that precede them in a questionnaire or interview schedule.
- Questions can be tested and improved through review by experts, focus group discussions, cognitive interviews, and pilot testing. Every questionnaire and interview schedule should be pretested on a small sample that is like the sample to be surveyed.
- The cover letter for a mailed questionnaire should be credible, personalized, interesting, and responsible.
- Response rates in mailed surveys are typically well below 70%, unless multiple mailings are made to nonrespondents and the questionnaire and cover letter are attractive, interesting, and carefully planned. Response rates for group-administered surveys are usually much higher than for mailed surveys.
- Phone interviews using random digit dialing (RDD) allow fast turnaround and efficient sampling. Multiple callbacks are often required, and the rate of nonresponse to phone interviews is rising. Phone interviews should be limited in length to about 30 to 45 minutes. In-person interviews have several advantages over other types of surveys: They allow longer and more complex interview schedules, monitoring of the conditions when the questions are answered, probing for respondents' understanding of the questions, and high response rates. However, the interviewer must balance the need to establish rapport with the respondent with the need to adhere to a standardized format.
- Electronic surveys may be e-mailed or posted on the web. Interactive voice

response systems using the telephone are another option. At this time, use of the Internet is not sufficiently widespread to allow e-mail or web surveys of the general population, but these approaches can be fast and efficient for populations with high rates of computer use.

- The decision to use a particular survey design must consider the unique features and goals of the study. In general, in-person interviews are the strongest but most expensive survey design.
- Most survey research poses few ethical problems because respondents can decline to participate—an option that should be stated clearly in the cover letter or introductory statement. Special care must be taken when questionnaires are administered in group settings (to “captive audiences”) and when sensitive personal questions are to be asked; subject confidentiality should always be preserved.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Response rates to phone surveys are declining, even as phone usage increases. Part of the problem is that lists of cell phone numbers are not available and wireless service providers do not allow outside access to their networks. Cell phone users may also have to pay for incoming calls. Do you think regulations should be passed to increase the ability of survey researchers to include cell phones in their random digit dialing surveys? How would you feel about receiving survey calls on your cell phone? What problems might result from "improving" phone survey capabilities in this way?
2. In-person interviews have for many years been the "gold standard" in survey research because the presence of an interviewer increases the response rate, allows better rapport with the interviewee, facilitates clarification of questions and instructions, and provides feedback about the interviewee's situation. However, researchers who design in-person interviewing projects are now increasingly using technology to ensure consistent questioning of respondents and to provide greater privacy while respondents are answering questions. But having a respondent answer questions on a laptop while the interviewer waits is a very different social process than asking the questions verbally. Which approach would you favor in survey research? What trade-offs can you suggest there might be in quality of information collected, rapport building, and interviewee satisfaction?

Finding Research

1. What resources are available for survey researchers? This question can be answered in part through careful inspection of a website maintained by the Survey Research Laboratory at the University of Illinois at Chicago (www.srl.uic.edu/srllink/srllink.htm#Organizations). Spend some time reviewing these resources, and write a brief summary of them.
2. Go to the Research Triangle Institute site at www.rti.org. Click on “Survey Research & Services” then “Innovations.” Read about their methods for computer-assisted interviewing and their cognitive laboratory methods for refining questions. What does this add to my treatment of these topics in this chapter? Give specific examples.

Critiquing Research

1. Read one of the original articles that reported one of the surveys described in this chapter. Critique the article using the questions presented in Exhibit 12.2 on page 294 as your guide but focus particular attention on sampling, measurement, and survey design.
2. Each of the following questions was used in a survey that we received at some time in the past. Evaluate each question and its response choices using the guidelines for question writing presented in this chapter. What errors do you find? Try to rewrite each question to avoid such errors and improve question wording.

    1. The first question in an *Info World* (computer publication) "product evaluation survey":

       How interested are you in PostScript Level 2 printers?

       ____Very ____Somewhat ____Not at all

    2. From the Greenpeace National Marine Mammal Survey:

       Do you support Greenpeace's nonviolent direct action to intercept whaling ships, tuna fleets, and other commercial fishermen in order to stop their wanton destruction of thousands of magnificent marine mammals?

       ____Yes ____No ____Undecided

    3. From a U.S. Department of Education survey of college faculty:

       How satisfied or dissatisfied are you with each of the following aspects of your instructional duties at this institution?

	Very Dissat.	Somewhat Dissat.	Somewhat Satisf.	Very Satisf.
i. The authority I have to make decisions about what courses I teach	1	2	3	4
ii. Time available for working with students as advisor, mentor	1	2	3	4

    d. From a survey about affordable housing in a Massachusetts community:

Higher than single-family density is acceptable to make housing affordable.

Strongly Agree	Undecided	Disagree	Strongly Agree	Disagree
1	2	3	4	5

e. From a survey of faculty experience with ethical problems in research:

Are you reasonably familiar with the codes of ethics of any of the following professional associations?

	Very Familiar	Familiar	Not Too Familiar
American Sociological Association	1	2	0
Society for the Study of Social Problems	1	2	0
American Society of Criminology	1	2	0

If you are familiar with any of the above codes of ethics, to what extent do you agree with them?

Strongly Agree _____ Agree _____ No opinion _____ Disagree _____ Strongly Disagree _____

Some researchers have avoided using a *professional code of ethics* as a guide for the following reason. Which responses, if any, best describe your reasons for not using all or any of parts of the codes?

	Yes	No
1. Vagueness	1	0
2. Political pressures	1	0
3. Codes protect only individuals, not groups	1	0

f. From a survey of faculty perceptions:

Of the students you have observed while teaching college courses, please indicate the percentage who significantly improve their performance in the following areas.

Reading ___%

Organization ___%
Abstraction ___%

g. From a University of Massachusetts, Boston, student survey:

A person has a responsibility to stop a friend or relative from driving when drunk.
Strongly Agree ____ Agree ____ Disagree ____ Strongly Disagree ____
Even if I wanted to, I would probably not be able to stop most people from driving drunk.
Strongly Agree ____ Agree ____ Disagree ____ Strongly Disagree ____

3. We received in a university mailbox some years ago a two-page questionnaire that began with the following “cover letter” at the top of the first page:

Critique this cover letter and then draft a more persuasive one.

4. Go to the UK Data Service at http://discover.ukdataservice.ac.uk/variables. In the search box, enter topics of interest such as “health” or “homelessness.” Review five questions for two topic areas and critique them in terms of the principles for question writing that you have learned. Do you find any question features that might be attributed to the use of British English?

Faculty Questionnaire

This survey seeks information on faculty perception of the learning process and student performance in their undergraduate careers. Surveys have been distributed in nine universities in the Northeast through random deposit in mailboxes of selected departments. This survey is being conducted by graduate students affiliated with the School of Education and the Sociology Department. We greatly appreciate your time and effort in helping us with our study.

Doing Research

1. Write 10 questions for a 1-page questionnaire that concerns a possible research question. Your questions should operationalize at least three of the variables on which you have focused, including at least one independent and one dependent variable. (You may have multiple questions to measure some variables.) Make all but one of your questions closed ended.
2. Conduct a preliminary pretest of the questionnaire by conducting cognitive interviews with two students or other persons like those to whom the survey is directed. Follow up the closed-ended questions with open-ended probes that ask the respondents what they meant by each response or what came to mind when they were asked each question. Take account of the feedback you receive when you revise your questions.
3. Polish the organization and layout of the questionnaire, following the guidelines in this chapter. Prepare a rationale for the order of questions in your questionnaire. Write a cover letter directed to the appropriate population that contains appropriate statements about research ethics (human subject issues).

Ethics Questions

1. Group-administered surveys are easier to conduct than other types of surveys, but they always raise an ethical dilemma. If a teacher allows a social research survey to be distributed in class, or if an employer allows employees to complete a survey on company time, is the survey truly voluntary? Is it sufficient to read a statement to the group stating that their participation is entirely up to them? How would you react to a survey in your class? What general guidelines should be followed in such situations?
2. Patricia Tjaden and Nancy Thoennes (2000) sampled adults with random digit dialing to study violent victimization from a nationally representative sample of adults. What ethical dilemmas do you see in reporting victimizations that are identified in a survey? What about when the survey respondents are under the age of 18? What about children under the age of 12?

Video Interview Questions

Listen to the researcher interview for Chapter 8 at edge.sagepub.com/chamblissmssw5e.

1. What two issues should survey researchers consider when designing questions?
2. Why is cognitive testing of questions important?

Chapter 8 Elementary Quantitative Data Analysis

©iStockphoto.com/PeskyMonkey

Learning Objectives

1. List the options for entering data for quantitative analysis.
2. Identify the types of graphs and statistics that are appropriate for analysis of variables at each level of measurement.
3. List the guidelines for constructing frequency distributions.
4. Discuss the advantages and disadvantages of using each of the three measures of central tendency.
5. Define the concept of skewness and explain how it can influence measures of central tendency.
6. Explain how to percentage a cross-tabulation table and how cross-tabulation can be used.
7. Discuss the reasons for conducting an elaboration analysis.
8. Know how to obtain secondary data.
9. Understand the concept and concerns in analyzing "big data."
10. Be aware of ethical guidelines for statistical analyses.

"Show me the data," says your boss. Presented with a research conclusion, most people—not just bosses—want evidence to support it; presented with piles of data, you the researcher need to uncover what it all means. To handle the data gathered by your research, you need to use straightforward methods of data analysis.]

Research|Social Impact Link

Read more on how quantitative data is impacting our daily lives.

In this chapter, we introduce several common statistics used in social research and explain how they can be used to make sense of the "raw" data gathered in your research. Such **quantitative data analysis**, using numbers to discover and describe patterns in your data, is the most elementary use of social statistics.

Quantitative data analysis: Statistical techniques used to describe and analyze variation in quantitative measures.

Why Do Statistics?

A **statistic**, in ordinary language usage, is a numerical description of a population, usually based on a sample of that population. (In the technical language of mathematics, a *parameter* describes a population, and a *statistic* specifically describes a sample.) Some statistics are useful for describing the results of measuring single variables or for constructing and evaluating multi-item scales. These statistics include frequency distributions, graphs, measures of central tendency and variation, and reliability tests. Other statistics are used primarily to describe the association among variables and to control for other variables, and thus, to enhance the causal validity of our conclusions. Cross-tabulation, for example, is one simple technique for measuring association and controlling other variables; it is introduced in this chapter. All of these statistics are termed **descriptive statistics** because they describe the distribution of and relationship among variables. Statisticians also use **inferential statistics** to estimate the degree of confidence that can be placed in generalizations from a sample to the population from which the sample was selected.

Case Study: The Likelihood of Voting

In this chapter, we use for examples some data from the 2012 General Social Survey (GSS) on voting and other forms of political participation. What influences the likelihood of voting? Prior research on voting in both national and local settings provides a great deal of support for one hypothesis: The likelihood of voting increases with social status (Milbrath & Goel 1977: 92–95; Salisbury 1975: 326; Verba & Nie 1972: 126). We will find out whether this hypothesis was supported in the 2012 GSS and examine some related issues.

The variables we use from the 2012 GSS are listed in Exhibit 8.1. We use these variables to illustrate particular statistics throughout this chapter.

Statistic: A numerical description of some feature of a variable or variables in a sample from a larger population.

Descriptive statistics: Statistics used to describe the distribution of and relationship among variables.

Inferential statistics: Statistics used to estimate how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

How to Prepare Data for Analysis

Our analysis of voting in this chapter is an example of what is called *secondary data analysis*. It is secondary because we received the data secondhand. A great many high-quality datasets are available for reanalysis from the Inter-university Consortium for Political and Social Research at the University of Michigan (1996), and many others can be obtained from the government, individual researchers, and other research organizations

Video Link

Watch a clip about how quantitative data can affect policy.

If you have conducted your own survey or experiment, your quantitative data must be prepared in a format suitable for computer entry. Questionnaires or other data entry forms can be designed to facilitate this process (Exhibit 8.2). Data from such a form can be entered online, directly into a database, or first on a paper form and then typed or even scanned into a computer database. Whatever data-entry method is used, the data must be checked carefully for errors—a process called **data cleaning**. Most survey research organizations now use a database management program to monitor data entry so that invalid codes can be corrected immediately. After data are entered, a computer program must be written to "define the data." A data-definition program identifies the variables that are coded in each column or range of columns, attaches meaningful labels to the codes, and distinguishes values representing missing data. The procedures vary depending on the specific statistical package used.

Exhibit 8.1 List of GSS 2010 Variables for Analysis of Voting

Variable[a]	SPSS Variable Name	Description
Family income	INCOME06	Family income (in categories)
	INCOMEFAM06	Family income (in approximate dollars)
Education	EDUCR	Years of education completed (6 categories)
	EDUC4	Years of education completed (4 categories)
	EDUC3	Years of education (3 categories)
Age	AGE4	Years old (4 categories)
	AGER	Years old (in decades)
Gender	SEX	Sex
Marital status	MARITAL	Married, never married, widowed, divorced
Race	RACED	White, minority
Politics	PARTYID3	Political party affiliation
Voting	VOTE08	Voted in 2008 presidential election (yes/no)
Political views	POLVIEW3	Liberal, moderate, conservative
Interpersonal trust	TRUSTD	Believe other people can be trusted

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Data cleaning: The process of checking data for errors after the data have been entered in a computer file.

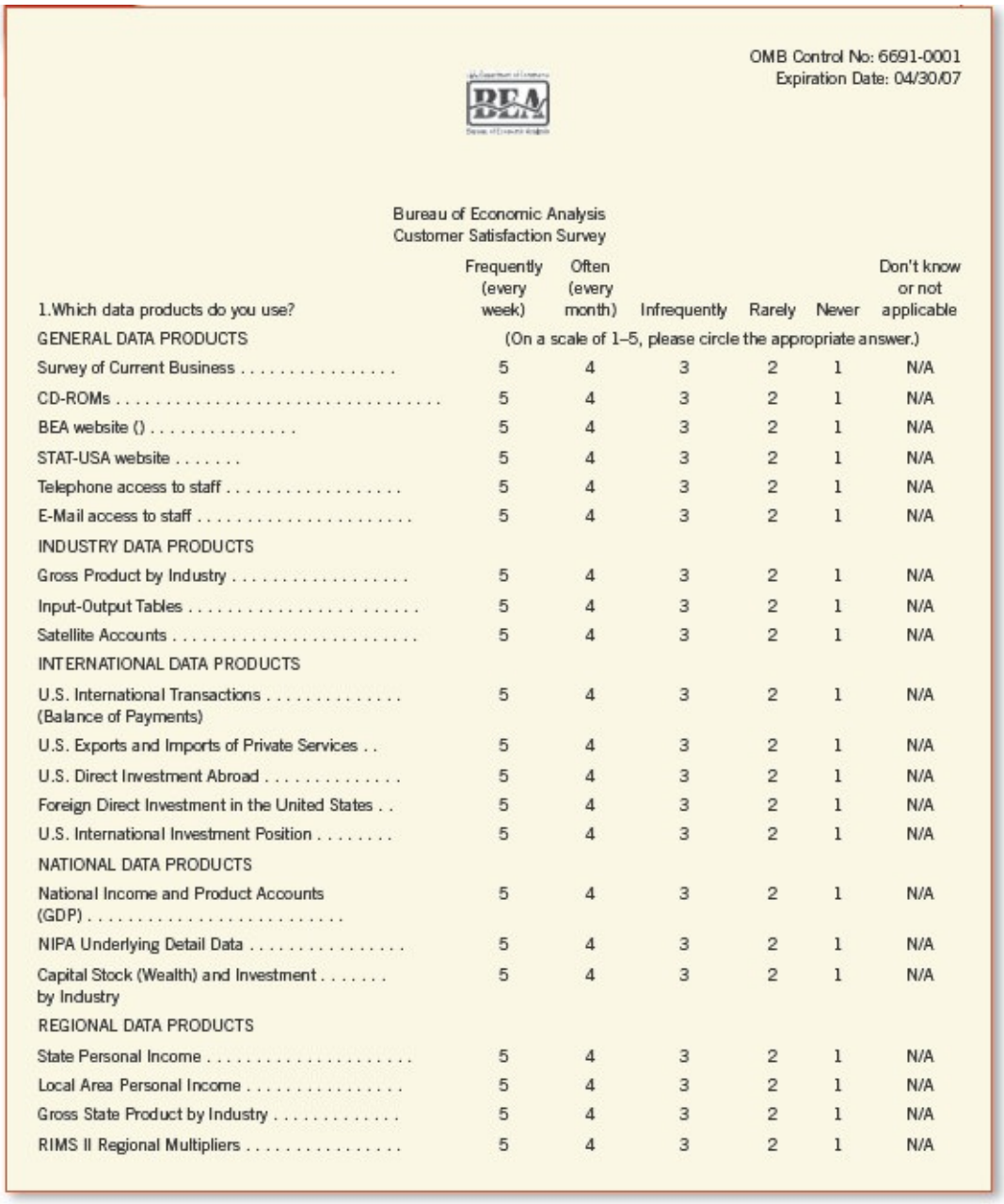
What Are the Options for Displaying Distributions?

The first step in data analysis is usually to discover the variation in each variable of interest. How many people in the sample are married? What is their typical income? Did most of them complete high school? Graphs and frequency distributions are the two most popular display formats. Whatever format is used, the primary concern of the analyst is to display accurately the distribution's shape—that is, to show how cases are distributed across the values of the variable.

Three features are important in describing the shape of the distribution: (1) **central tendency**, (2) **variability**, and (3) **skewness** (lack of symmetry). These three features can be represented in a graph or in a frequency distribution.

We now examine graphs and frequency distributions that illustrate the three features of shape. Several summary statistics used to measure specific aspects of central tendency and variability are presented in a separate section.

Exhibit 8.2 Online Data Collection Form



OMB Control No: 6691-0001
Expiration Date: 04/30/07

Bureau of Economic Analysis
Customer Satisfaction Survey

1.Which data products do you use?	Frequently (every week)	Often (every month)	Infrequently	Rarely	Never	Don't know or not applicable
GENERAL DATA PRODUCTS	(On a scale of 1–5, please circle the appropriate answer.)					
Survey of Current Business	5	4	3	2	1	N/A
CD-ROMs	5	4	3	2	1	N/A
BEA website ()	5	4	3	2	1	N/A
STAT-USA website	5	4	3	2	1	N/A
Telephone access to staff	5	4	3	2	1	N/A
E-Mail access to staff	5	4	3	2	1	N/A
INDUSTRY DATA PRODUCTS						
Gross Product by Industry	5	4	3	2	1	N/A
Input-Output Tables	5	4	3	2	1	N/A
Satellite Accounts	5	4	3	2	1	N/A
INTERNATIONAL DATA PRODUCTS						
U.S. International Transactions (Balance of Payments)	5	4	3	2	1	N/A
U.S. Exports and Imports of Private Services	5	4	3	2	1	N/A
U.S. Direct Investment Abroad	5	4	3	2	1	N/A
Foreign Direct Investment in the United States	5	4	3	2	1	N/A
U.S. International Investment Position	5	4	3	2	1	N/A
NATIONAL DATA PRODUCTS						
National Income and Product Accounts (GDP)	5	4	3	2	1	N/A
NIPA Underlying Detail Data	5	4	3	2	1	N/A
Capital Stock (Wealth) and Investment by Industry	5	4	3	2	1	N/A
REGIONAL DATA PRODUCTS						
State Personal Income	5	4	3	2	1	N/A
Local Area Personal Income	5	4	3	2	1	N/A
Gross State Product by Industry	5	4	3	2	1	N/A
RIMS II Regional Multipliers	5	4	3	2	1	N/A

Source: U.S. Bureau of Economic Analysis, Communications Division. 2004. Customer satisfaction survey report, FY 2004. Washington, DC: U.S. Department of Commerce, 14. From http:// www.bea.gov/ bea/about/cssr_2004_complete.pdf (accessed September 28, 2008).

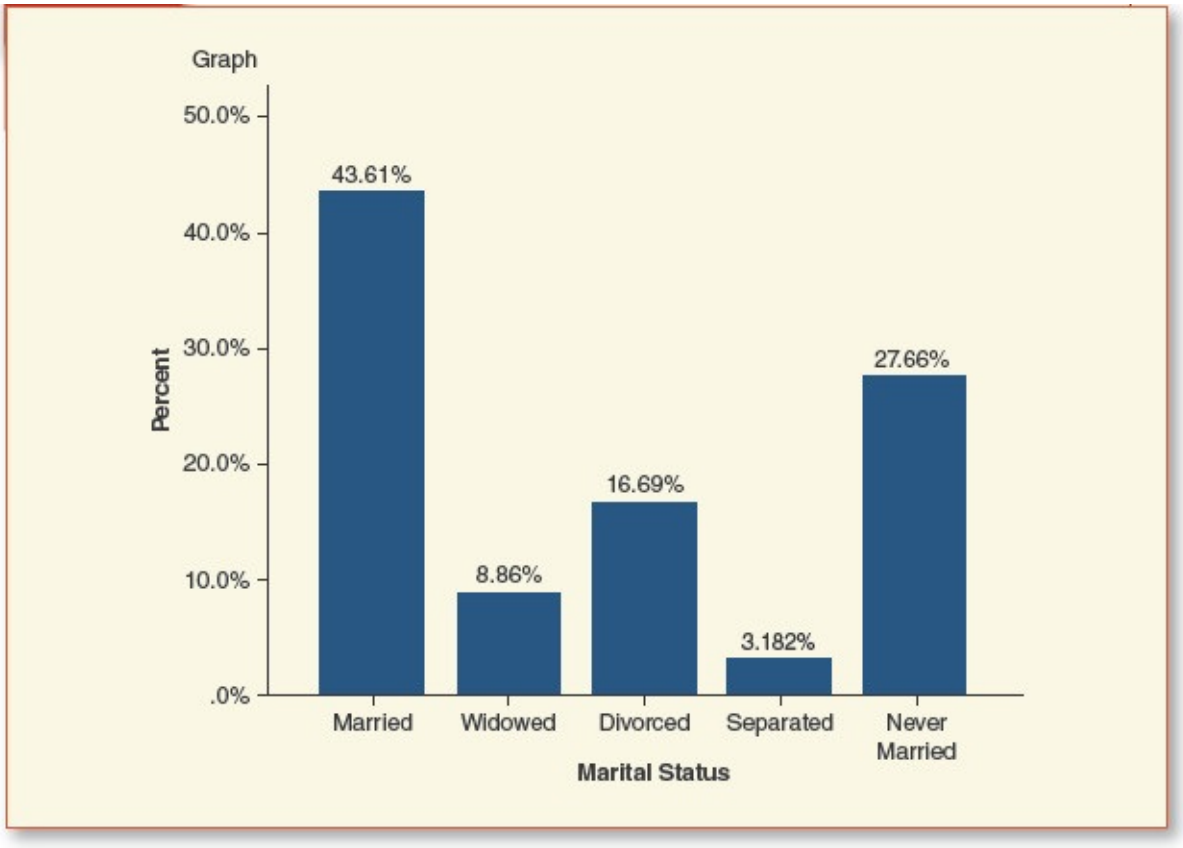
Graphs

There are many types of graphs, but the most common and most useful for the statistician are bar charts, histograms, and frequency polygons. Each has two axes, the vertical axis (the *y*-axis) and the horizontal axis (the *x*-axis), and labels to identify the variables and the values, with tick marks showing where each indicated value falls along each axis.

A **bar chart** contains solid bars separated by spaces. It is a good tool for displaying the distribution of variables measured in discrete categories (e.g., nominal variables such as religion or marital status) because such categories don't blend into each other. The bar chart of marital status in Exhibit 8.3 indicates that about half of adult Americans were married at the time of the survey. Smaller percentages were divorced, separated, widowed, or never married. The most common value in the distribution is *married.* There is a moderate amount of variability in the distribution because the half that is not married is spread across the categories of widowed, divorced, separated, and never married. Because marital status is not a quantitative variable, the order in which the categories are presented is arbitrary, and there is no need to discuss skewness.

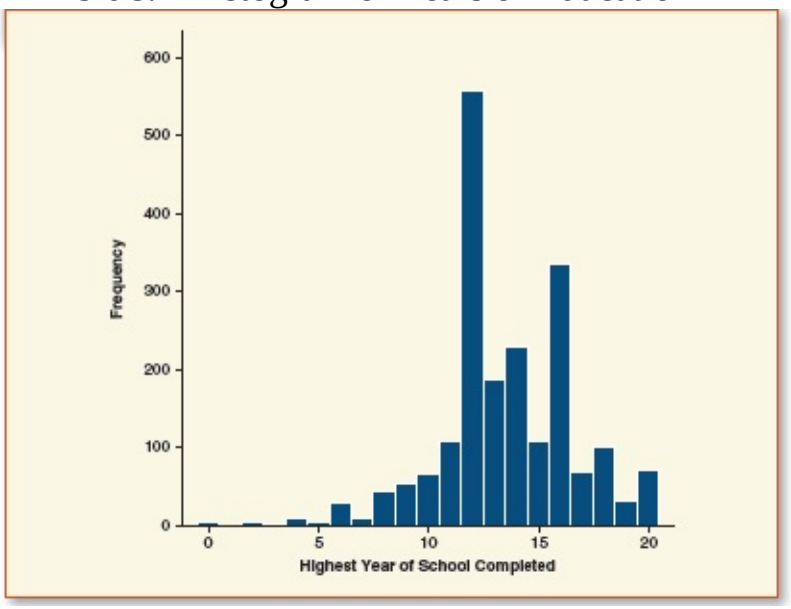
Histograms, in which the bars are adjacent, are used to display the distribution of quantitative variables that vary along a continuum that has no necessary gaps. Exhibit 8.4 shows a histogram of years of education from the 2012 GSS data. The distribution has a clump of cases centered at 12 years. The distribution is skewed because there are more cases just above the central point than below it.

Exhibit 8.3 Bar Chart of Marital Status



Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Exhibit 8.4 Histogram of Years of Education



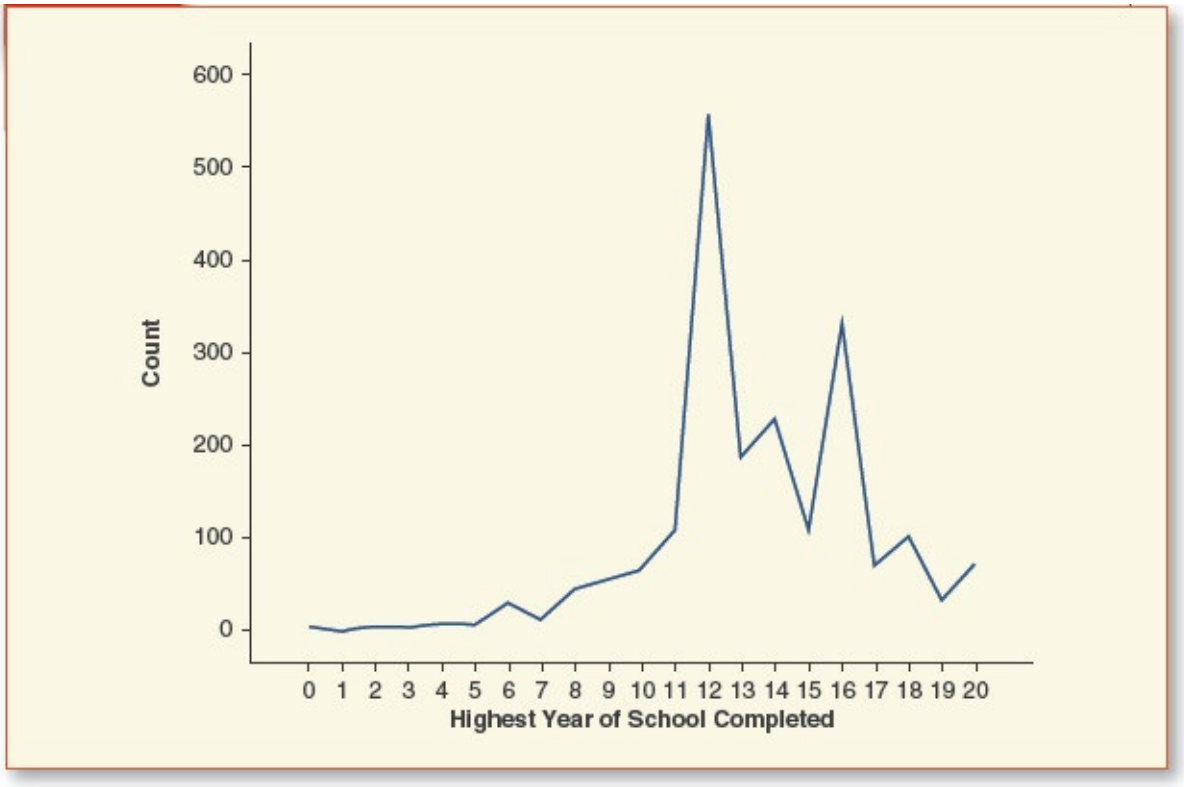
Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

In a **frequency polygon**, a continuous line connects the points representing the number or percentage of cases with each value. It is easy to see in the frequency polygon of years of education in Exhibit 8.5 that the most common value is 12 years (high school completion) and that this value seems to be the center of the distribution. There is moderate variability in the distribution, with many cases having more than 12 years of education and almost one third having completed at least 4 years of college (16 years). The distribution is highly skewed in the negative direction, with few respondents reporting less than 10 years of education.

If graphs are misused, they can distort rather than display the shape of a distribution. Compare, for example, the two graphs in Exhibit 8.6. The first graph shows that high school seniors reported relatively stable rates of lifetime use of cocaine between 1980 and 1985.

The second graph, using exactly the same numbers, appeared in a 1986 *Newsweek* article on "the coke plague" (Orcutt & Turner 1993). To look at this graph, you would think that the rate of cocaine usage among high school seniors had increased dramatically during this period. But the difference between the two graphs actually results simply from changes in how the graphs were drawn. In the *Newsweek* graph, the percentage scale on the vertical axis begins at 15 rather than at 0, making what was about a 1 percentage point increase look very big indeed. In addition, omission from this graph of the more rapid increase in reported usage between 1975 and 1980 makes it look as if the tiny increase in 1985 were a new, and thus more newsworthy, crisis. Finally, these numbers report "lifetime use," not current or recent use; such numbers can drop only when anyone who has used cocaine dies. The graph is, in total, grossly misleading.

Exhibit 8.5 Frequency Polygon of Years of Education



Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Adherence to several guidelines (Tufte 1983; Wallgren et al. 1996) will help you spot such problems and avoid them in your own work:

- Begin the graph of a quantitative variable at 0 on both axes. The difference between bars can be misleadingly exaggerated by cutting off the bottom of the vertical axis and displaying less than the full height of the bars. It may at times be reasonable to violate this guideline, as when an age distribution is presented for a sample of adults; but in this case, be sure to mark the break clearly on the axis.
- Always use bars of equal width. Bars of unequal width, including pictures instead of bars, can make particular values look as if they carry more weight than their frequency warrants.
- Ensure that the two axes, usually, are of approximately equal length. Either shortening or lengthening the vertical axis will obscure or accentuate the differences in the number of cases between values.
- Avoid “chart junk”—a lot of verbiage or excessive marks, lines, lots of

cross-hatching, and the like. It can confuse the reader and obscure the shape of the distribution.

Central tendency: The most common value (for variables measured at the nominal level) or the value around which cases tend to center (for a quantitative variable).

Variability: The extent to which cases are spread out through the distribution or clustered around just one value.

Skewness: The extent to which cases are clustered more at one or the other end of the distribution of a quantitative variable rather than in a symmetric pattern around its center. Skew can be positive (a right skew), with the number of cases tapering off in the positive direction, or negative (a left skew), with the number of cases tapering off in the negative direction.

Bar chart: A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces.

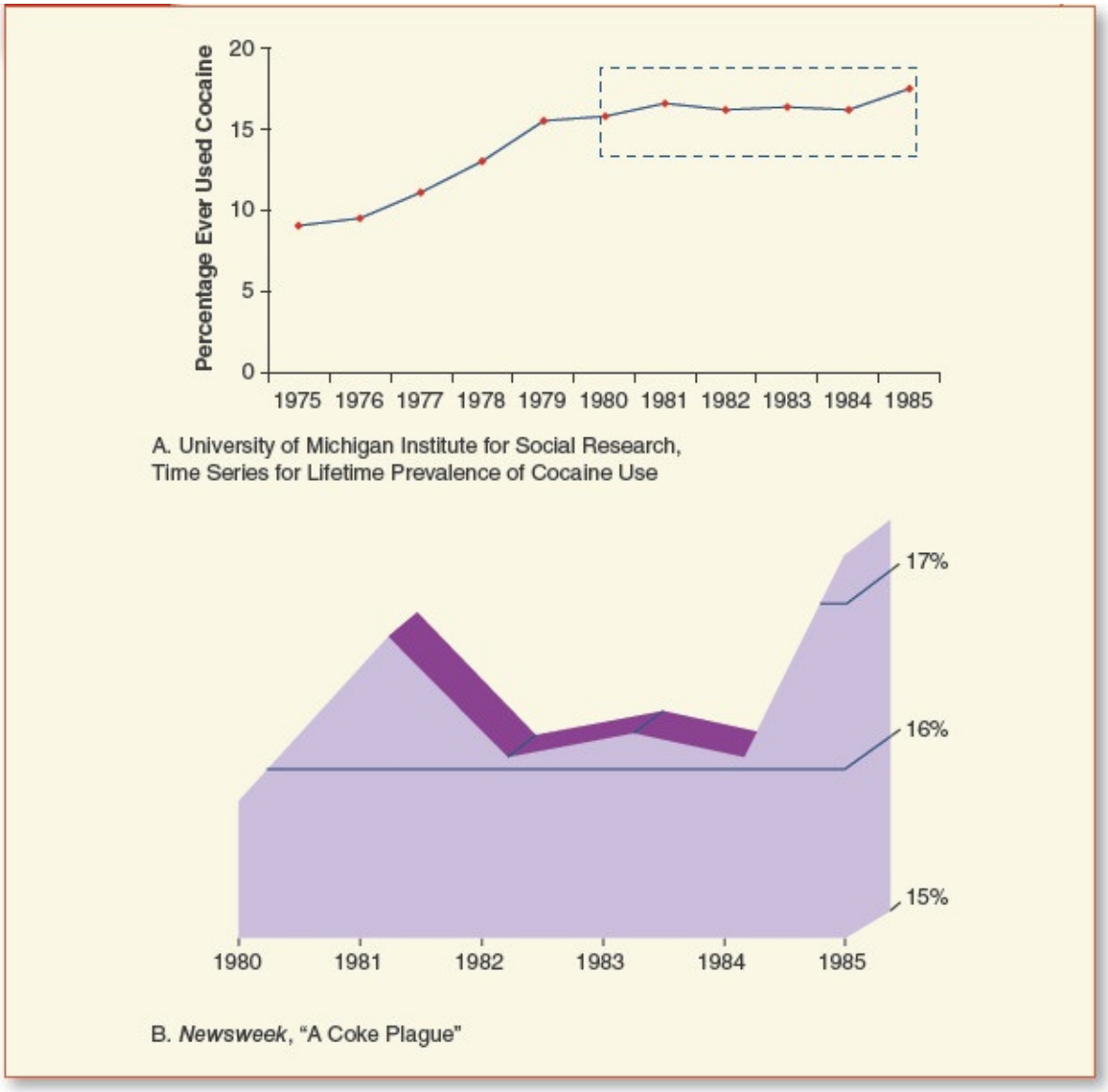
Histogram: A graphic for quantitative variables in which the variable's distribution is displayed with adjacent bars.

Frequency polygon: A graphic for quantitative variables in which a continuous line connects data points representing the variable's distribution.

Frequency Distributions

Another good way to present a univariate (one-variable) distribution is with a **frequency distribution**. A frequency distribution displays the number, **percentage** (the relative frequencies), or both corresponding to each of a variable's values. A frequency distribution will usually be labeled with a title, a stub (labels for the values), a caption, and perhaps the number of missing cases. If percentages are presented rather than frequencies (sometimes both are included), the total number of cases in the distribution (the **base number *N***) should be indicated (Exhibit 8.7).

Exhibit 8.6 Two Graphs of Cocaine Usage



Source: Adapted from Orcutt, James D., and J. Blake Turner. 1993. Shocking numbers and graphic accounts: Quantified images of drug problems in the print media. Social Problems 49: 190–206. Copyright 1993 by the Society for the Study of Social Problems. Reprinted by permission.

Constructing and reading frequency distributions for variables with few values is not difficult. The frequency distribution of voting in Exhibit 8.7, for example, shows that 72.9% of the respondents eligible to vote said they voted and that 27.1% reported they did not vote. The total number of respondents to this question was 1,974, although 1,789 actually were interviewed. The rest were

ineligible to vote, just refused to answer the question, said they did not know whether they had voted, or gave no answer.

When the distributions of variables with many values (for instance, age) are to be presented, the values must first be grouped. Exhibit 8.8 shows both an ungrouped and a grouped frequency distribution of age. You can see why it is so important to group the values, but we have to be sure that in doing so, we do not distort the distribution. Follow these two rules, and you'll avoid problems:

1. Categories should be logically defensible and should preserve the shape of the distribution.
2. Categories should be mutually exclusive and exhaustive so that every case is classifiable in one and only one category.

Exhibit 8.7 Frequency Distribution of Voting in the 2008 Presidential Election

Value	Frequency	Valid Percentage
Voted	1304	72.9%
Did not vote	485	27.1%
Ineligible	159	
Don't know	22	
No answer	4	
Total %		100.0%
N	1974	(1789)

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Frequency distribution: Numerical display showing the number of cases, and usually the percentage of cases (the relative frequencies), corresponding to each value or group of values of a variable.

Percentage: The relative frequency, computed by dividing the frequency of cases in a particular category by the total number of cases and multiplying by 100.

Base number (*N*): The total number of cases in a distribution.

What Are the Options for Summarizing Distributions?

Summary statistics describe particular features of a distribution and facilitate comparison among distributions. We can, for instance, show that average income is higher in Connecticut than in Mississippi and higher in New York than in Louisiana. But if we just use one number to represent a distribution, we lose information about other aspects of the distribution's shape. For example, a measure of central tendency (such as the mean or average) would miss the point entirely for an analysis about differences in income inequality among states. A high average income could as easily be found in a state with little income inequality as in one with much income inequality; the average says nothing about the distribution of incomes. For this reason, analysts who report summary measures of central tendency usually also report a summary measure of variability or present the distributions themselves to indicate skewness.

Video Link

Watch a discussion about why data visualization is important.

Exhibit 8.8 Grouped Versus Ungrouped Frequency Distributions

Ungrouped		Grouped	
Age	Percentage	Age	Percentage
18	0.6%	18–19	1.6%
19	1.0%	20–29	15.2%
20	1.0%	30–39	19.8%
21	1.6%	40–49	18.1%
22	1.6%	50–59	17.1%
23	1.5%	60–69	14.5%
24	1.4%	70–79	8.7%
25	1.6%	80–89	5.0%
26	1.5%		100.0% (1969)
27	1.6%		
28	2.1%		
29	1.4%		
30	2.4%		
31	1.8%		
32	2.4%		
33	2.0%		
34	1.7%		
35	1.8%		
36	1.7%		
37	2.0%		
38	1.9%		
39	2.1%		
40	1.5%		
41	2.0%		
42	2.2%		
43	1.5%		
44	1.5%		
45	1.8%		
46	1.7%		
...	...		

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Measures of Central Tendency

Central tendency is usually summarized with one of three statistics: the mode, the median, or the mean. For any particular application, one of these statistics may be preferable, but each has a role to play in data analysis. To choose an appropriate measure of central tendency, the analyst must consider a variable's level of measurement, the skewness of a quantitative variable's distribution, and the purpose for which the statistic is used.

Encyclopedia Link

Read about when to use measures of central tendency.

Mode

The **mode** is the most frequent value in a distribution. In a distribution of Americans' religious affiliations, Protestant Christian is the most frequently occurring value—the largest single group. In an age distribution of college students, 18- to 22-year-olds are by far the largest group and, therefore, the mode. One silly, but easy, way to remember the definition of the *mode* is to think of apple pie *á la mode*, which means pie with a big blob of vanilla ice cream on top. Just remember, the mode is where the big blob is—the largest collection of cases.

The mode is also sometimes termed the **probability average**, because being the most frequent value, it is the most probable. For example, if you were to pick a case at random from the distribution of age (Exhibit 8.8), the probability of the case being in his or her 30s would be 19.8%—the most probable value in the distribution.

The mode is used much less often than the other two measures of central tendency because it can so easily give a misleading impression of a distribution's central tendency. One problem with the mode occurs when a distribution is **bimodal**, in contrast to being **unimodal**. A bimodal distribution has two categories with a roughly equal number of cases and clearly more cases than the

other categories. In this situation, there is no single mode.

Nevertheless, there are occasions when the mode is very appropriate. The mode is the only measure of central tendency that can be used to characterize the central tendency of variables measured at the nominal level. In addition, because it is the most probable value, it can be used to answer questions such as which ethnic group is most common in a given school.

Research in the News

General Social Survey Shows Infidelity on the Rise

Since 1972, about 12% of married men and 7% of married women have said each year that they have had sex outside their marriage. However, the lifetime rate of infidelity for men older than age 60 increased from 20% in 1991 to 28% in 2006, whereas for women in this age group, it increased from 5% to 15%. Infidelity has also increased among those younger than age 35: from 15% to 20% among young married men and from 12% to 15% among young married women. Conversely, couples appear to be spending slightly more time with each other.

1. 1.Do you think that these changes reflect shifts in morals or other types of changes? What other variables would you want to include in an analysis to test alternative explanations for the change?
2. 2.What would you want to measure about the characteristics of the interview situation it

News Source: Parker-Pope, Tara. 2008. Love, sex, and the changing landscape of infidelity. *New York Times*, October

Mode (probability average): The most frequent value in a distribution; also termed the *probability average*.

Bimodal: A distribution in which two nonadjacent categories have about the same number of cases and these categories have more cases than any others.

Unimodal: A distribution of a variable in which only one value is the most frequent.

Median: The position average, or the point, that divides a distribution in half (the 50th percentile).

Median

The **median** is the position average, or the point that divides the distribution in half (the 50th percentile). Think of the median of a highway—it divides the road exactly in two parts. To determine the median, we simply array a distribution's values in numerical order and find the value of the case that has an equal number of cases above and below it. If the median point falls between two cases (which happens if the distribution has an even number of cases), the median is defined as the average of the two middle values and is computed by adding the values of

the two middle cases and dividing by 2. The median is not appropriate for variables that are measured at the nominal level; their values cannot be put in order, so there is no meaningful middle position.

Exhibit 8.9 Years of Education Completed

Years of Education	Percentage
Less than 8	3.2%
8–11	12.9%
12	27.4%
13–15	26.5%
16	15.6%
17 or more	14.4%
	100.0 (1972)

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

The median in a frequency distribution is determined by identifying the value corresponding to a cumulative percentage of 50. Starting at the top of the years of education distribution in Exhibit 8.9, for example, and adding the percentages, we find that we reach 43.5% in the 12-years category and then 69.8% in the 13- to 15-years category. The median is therefore 13 to 15.

Mean

The **mean** is just the arithmetic average. (Many people, you'll notice, use the word *average* a bit more generally to designate everything we've called central tendency.) In calculating a mean, any higher numbers pull it up, and any lower numbers pull it down. Therefore, it accounts for the values of each case in a distribution—it is a weighted average. (The median, by contrast, only depends on whether the numbers are higher or lower compared with the middle, not *how* high or low.)

The mean is computed by adding up the values of all the cases and dividing the result by the total number of cases.
Mean=Sumofvaluesofcases/Numberofcases
Mean = Sum of value of cases / Number of cases

In algebraic notation, the equation is $\bar{X}=\sum x/N$ $\bar{X} = \Sigma x/N$. For example, to calculate the mean value of eight cases, we add the values of all the cases ($\sum x_i$ and divide by the number of cases (N):
(28+117+42+10+77+51+64+55)/8=55.5
$(28 + 117 + 42 + 10 + 77 + 51 + 64 + 55) / 8 = 55.5$

Computing the mean requires adding the values of the cases. So it makes sense to compute a mean only if the values of the cases can be treated as actual quantities—that is, if they reflect an interval or ratio level of measurement—or if we assume that an ordinal measure can be treated as an interval (which is a fairly common practice). It makes no sense to calculate the mean of a qualitative (nominal) variable such as religion, for example. Imagine a group of four people in which there were two Protestants, one Catholic, and one Jew. To calculate the mean, you would need to solve the equation (Protestant + Protestant + Catholic + Jew) / 4 = *?* Even if you decide that Protestant = 1, Catholic = 2, and Jew = 3 for data entry purposes, it still doesn't make sense to add these numbers because they don't represent quantities of religion. In general, certain statistics (such as the mean) can apply only if there is a high enough level of measurement.

Mean: The arithmetic, or weighted, average computed by adding the value of all the cases and dividing by the total number of cases.

Median or Mean?

The mean is based on adding the value of all the cases, so it will be pulled in the direction of exceptionally high (or low) values. In a positively skewed distribution, the value of the mean is larger than the median—more so the more extreme the skew. For instance, in Seattle, the presence of Microsoft co-founder Bill Gates—possibly the world's richest person—probably pulls the mean wealth number up quite a bit. One extreme case can have a disproportionate effect on the mean.

Research|Social Impact Link

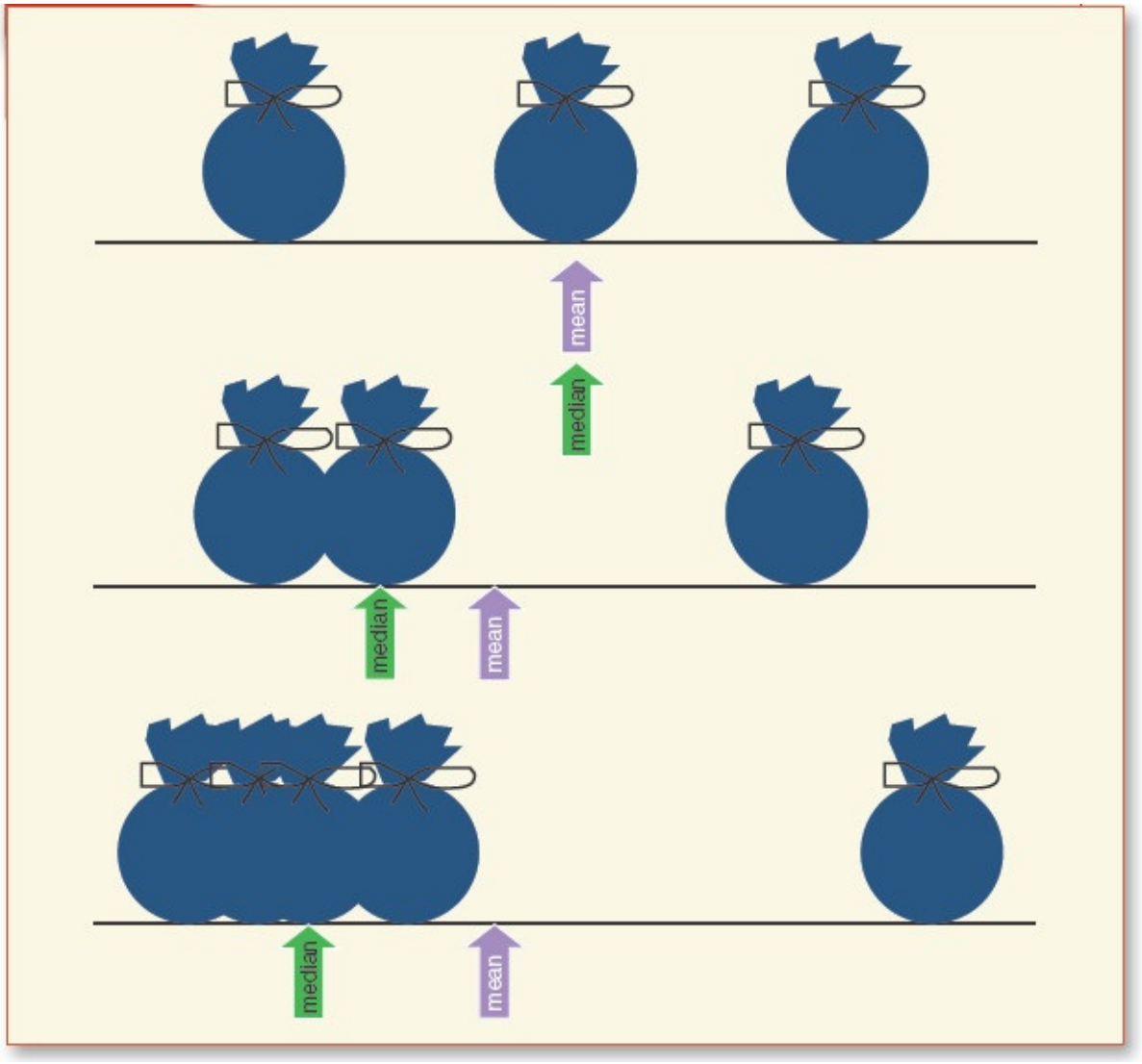
Read about the difference in measures of central tendancy.

This differential impact of skewness on the median and mean is illustrated in Exhibit 8.10. On the first balance beam, the cases (bags) are spread out equally, and the median and mean are in the same location. On the second balance beam, the median corresponds to the value of the middle case, but the mean is pulled slightly upward toward the value of the one case with an unusually high value. On the third beam, the mean is clearly pulled up toward an unusual value. In some distributions, the two measures will have markedly different values, and in such instances, usually the median is preferred. (Income is a very common variable that is best measured by the median, for instance.)

Measures of Variation

Central tendency is only one aspect of the shape of a distribution—the most important aspect for many purposes but still just a piece of the total picture. The distribution, we have seen, also matters. It is important to know that the median household income in the United States is a bit over $50,000 a year, but if the variation in income isn't known—the fact that incomes range from zero to hundreds of millions of dollars—we haven't really learned much. Measures of variation capture how widely and densely spread income (for instance) is. Four popular measures of variation for quantitative variables are the range, the interquartile range, the variance, and the standard deviation (which is the single most popular measure of variability). Each conveys a certain kind of information, with strengths and weaknesses. Statistical measures of variation are used infrequently with qualitative variables and are not presented here.

Exhibit 8.10 The Mean as a Balance Point



Range

The **range** is the simplest measure of variation, calculated as the highest value in a distribution minus the lowest value, plus 1:

$$\text{Range} = \text{Hightestvalue} - \text{Lowestvalue} + 1$$

Range = Highest value – Lowest value + 1

It often is important to report the range of a distribution—to identify the whole range of possible values that might be encountered. However, because the range can be altered drastically by just one exceptionally high or low value—termed an **outlier**—it's not a good summary measure for most purposes.

Research That Matters

Why do some urban youth grow up to become regular voters in elections, but others do not? Could rates of voting be improved with education programs targeted to high school students? Alison K. Cohen and Benjamin W. Chaffee investigated the first question, hoping to help design programs that would answer the second.

At the beginning of the school year, they collected survey data from youth in Providence, Rhode Island, and Boston, Massachusetts. Questionnaires were distributed by classroom teachers and anonymously completed by students who signed an informed consent form. Questions were designed to measure civic knowledge, attitudes, and behaviors, as well as likelihood of voting and various academic and social characteristics. Overall, Cohen and Chaffee found that the more that students knew about civic affairs, the more likely they were to vote, but it wasn't so clear what educational programs would be most effective in increasing voting.

Source: Adapted from Cohen, Alison K., and Benjamin W. Chaffee. 2012. The relationship between adolescents' civic knowledge, civic attitude, and civic behavior and their self-reported future likelihood of voting. *Education, Citizenship and Social Justice* 8(1): 43–57.

Range: The true upper limit in a distribution minus the true lower limit (or the highest rounded value minus the lowest rounded value, plus 1).

Outlier: An exceptionally high or low value in a distribution.

Interquartile range: The range in a distribution between the end of the 1st quartile and the beginning of the 3rd quartile.

Quartiles: The points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases.

Variance: A statistic that measures the variability of a distribution as the average squared deviation of each case from the mean.

Interquartile Range

The **interquartile range** avoids the problem outliers create by showing the range where most cases lie. **Quartiles** are the points in a distribution that correspond to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases. You already know how to determine the 2nd quartile, corresponding to the point in the distribution covering half of the cases—it is another name for the median. The interquartile range is the difference between the 1st quartile and the 3rd quartile (plus 1).

Variance

Variance, in its statistical definition, is the average squared deviation of each case from the mean; you take each case's distance from the mean, square that number, and take the average of all such numbers. Thus, variance considers the amount by which each case differs from the mean. The variance is mainly useful for computing the standard deviation, which comes next in our list here. An example of how to calculate the variance, using the following formula, appears in Exhibit 8.11:

σ2=∑(Yi−Y¯i)2N

$$\sigma^2 = \frac{\sum (Y_i - \bar{Y}_i)^2}{N}$$

Exhibit 8.11 Calculation of the Variance

Case #	Score (X_i)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	21	–3.27	10.69
2	30	5.73	32.83
3	15	–9.27	85.93
4	18	–6.27	39.31
5	25	0.73	0.53
6	32	7.73	59.75
7	19	–5.27	27.77
8	21	–3.27	10.69
9	23	–1.27	1.61
10	37	12.73	162.05
11	26	1.73	2.99

Mean: $\bar{X}$ = 267/11 = 24.27

Sum of squared deviations = 434.15

Variance: σ2 = 434.15/11 = 39.47

Symbol key: $\bar{X}$ = mean; N = number of cases; S = sum over all cases; X_i = value of case i on variable X.

The variance is used in many other statistics, although it is more conventional to measure variability with the closely related standard deviation than with the variance.

Standard Deviation

Very roughly, the **standard deviation** is the distance from the mean that covers a clear majority of cases (about two thirds). More precisely, the standard deviation is simply the square root of the variance. It is the square root of the average squared deviation of each case from the mean:

$$\sigma = \sum(Yi - \bar{Y}i)N$$

$$\sigma = \sqrt{\frac{\sum(Y_i - \bar{Y_i})}{N}}$$

Symbol key: $\bar{X}$ = mean; N = number of cases; S = sum over all cases; X_i = value of case on i variable X; √ = square root.

The standard deviation has mathematical properties that make it the preferred measure of variability in many cases, particularly when a variable is normally distributed. A graph of a **normal distribution** looks like a bell, with one “hump” in the middle, centered around the population mean, and the number of cases tapering off on both sides of the mean (Exhibit 8.12). A normal distribution is symmetric: If you were to fold the distribution in half at its center (at the population mean), the two halves would match perfectly. If a variable is normally distributed, 68% of the cases (almost exactly two-thirds) will lie between ±1 standard deviation from the distribution’s mean, and 95% of the cases will lie between 1.96 standard deviations above and below the mean.

So the standard deviation, in a single number, tells you quickly about how wide the variation is of any set of cases, or the range in which most cases will fall. It’s very useful.

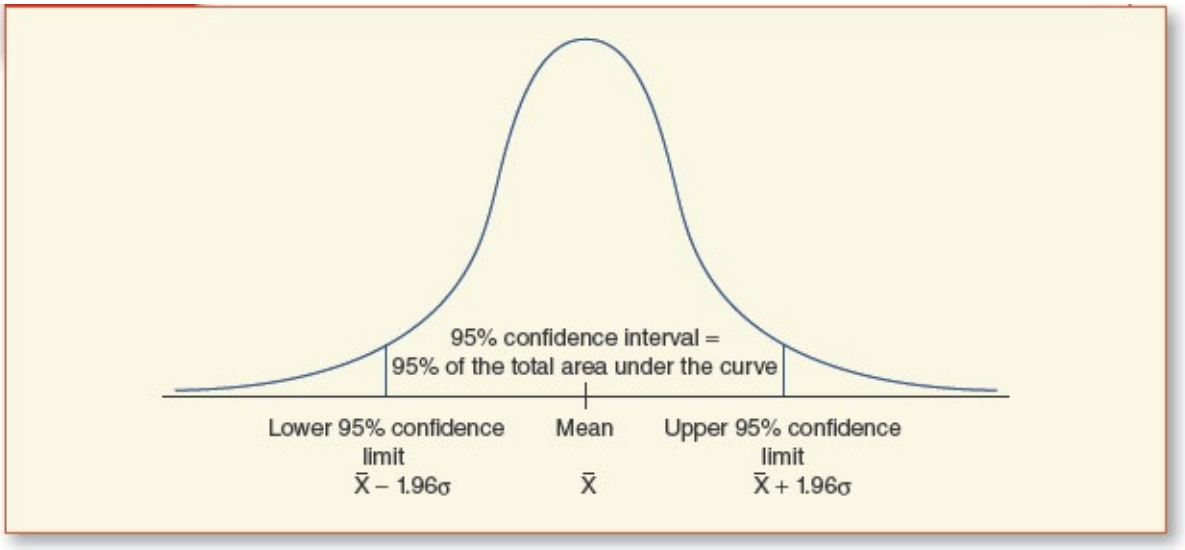
Standard deviation: The square root of the average squared deviation of each case from the mean.

Normal distribution: A symmetric distribution shaped like a bell and centered around the population mean, with the number of cases tapering off in a predictable pattern on both sides of the mean.

How Can We Tell Whether Two Variables Are Related?

Univariate distributions are nice, but they don't say how variables relate to each other—for instance, if religion affects education or if marital status is related to income. To establish cause, of course, one's first task is to show an association between independent and dependent variables (cause and effect). **Cross-tabulation** is a simple, easily understandable first step in such quantitative data analysis. Cross-tabulation displays the distribution of one variable within each category of another variable; it can also be termed a *bivariate distribution* because it shows two variables at the same time. Exhibit 8.13 displays the cross-tabulation of voting by income so that we can see if the likelihood of voting increases as income goes up.

Exhibit 8.12 The Normal Distribution



The "crosstab" table is presented first (the upper part) with frequencies and then again (the lower part) with percentages. The *cells* of the table are where row and column values intersect; for instance, the first cell is where < $20,000 meets Voted; 202 is the value. Each cell represents cases with a unique combination of values of the two variables. The independent variable is usually the column variable, listed across the top; the dependent variable, then, is usually the row variable. This format isn't necessary, but social scientists typically use it.

Interactive Exercises

Quantitative Data Analysis

Reading the Table

The first (upper) table in Exhibit 8.13 shows the raw number of cases with each combination of values of voting and family income. It is hard to look at the table in this form and determine whether there is a relationship between the two variables. What we really want to know is the likelihood, for any level of income, that someone voted. So we need to convert the cell frequencies into percentages. Percentages show the likelihood per 100 (*per cent* in Latin) that something occurs. The second table, then, presents the data as percentages within the categories of the independent variable (the column variable, in this case). In other words, the cell frequencies have been converted into percentages of the column totals (the *N* in each column). For example, in Exhibit 8.13, the number of people earning less than $20,000 who voted is 202 out of 362, or 55.8%. Because the cell frequencies have been converted to percentages of the column totals, the numbers total 100 in each column but not across the rows.

Exhibit 8.13 Cross-Tabulation of Voting in 2008 by Family Income: Cell Counts and Percentages

	Family Income			
Voting	<$20,000	$20,000–$39,999	$40,000–$74,999	$75,000+
	Cell Counts			
Voted	202	269	297	411
Did not vote	160	113	101	61
Total (n)	(362)	(382)	(398)	(472)
	Percentages			
Voted	55.8	70.4	74.6	87.1
Did not vote	44.2	29.6	25.4	12.9
Total	100	100	100	100

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Note carefully: You must *always* calculate percentages within levels of the independent variable—adding numbers down the columns in our standard format. In this example, we want to know the chance that a person with an income of less than $20,000 voted, so we calculate what percentage of those people voted. Then we *compare* that to the chance that people of other income levels voted. Calculating percentages across the table, by contrast, will not show the effect of the independent variable on voting. To repeat, *always* calculate percentages within levels of the independent variable (think: with**in** the **in**dependent variable).

To read the percentage table, compare the percentage distribution of voting/not voting across the columns. Start with the lowest income category (in the left column). Move slowly from left to right, looking at each distribution down the columns. As income increases, you will see that the percentage who voted also increases, from 55.8% of those with annual incomes under $20,000 (in the first cell in the first column) to 87.1% of those with incomes of $75,000 or more (the last cell in the body of the table in the first row). This result is consistent with the hypothesis: It seems that higher income is moderately associated with a greater likelihood of voting.

Now look at Exhibit 8.14, which relates gender (as the independent variable) to voting (the dependent variable). The independent variable is listed across the top, and the percentages have been calculated, correctly, down the columns with values of the independent variable. Does gender affect voting? As you look down the first column, you see that 70.7% of men voted; then, in the second column, 74.7% of women voted. Gender did, in this table, have some effect on voting. Women were more likely to vote.

Some standard practices should be followed in formatting percentage tables (crosstabs): When a table is converted to percentages, usually just the percentages in each cell should be presented, and not the number of cases in each cell. Include 100% at the bottom of each column (if the independent variable is the column variable) to indicate that the percentages add up to 100, as well as the base number (*N*) for each column (in parentheses). If the percentages add up to 99 or 101 because of rounding error, just indicate so in a footnote. As noted already, there is no requirement that the independent variable always be the column variable, although consistency within a report or paper is a must. If the independent variable is the row variable, we calculate percentages in the cells of the table on the row totals (the *N* in each row), and the percentages add

up to 100 across the rows.

Exhibit 8.15 shows two different tables. The top half shows voting by education —that is, the likelihood that a person with a given level of education voted in 2008. Look first at the voting distribution for high school graduates: The percent voting has jumped to more than 69%—a significant change from the percentage for grade school completers. As you move across to the numbers for some college, then college graduates, it becomes obvious that education has a major effect on a person's likelihood of voting.

Exhibit 8.14 Voting in 2008 by Gender

	Gender	
Voting	**Male**	**Female**
Voted	70.7%	74.7%
Did not vote	29.3%	25.3%
Total	100.0%	100.0%
(n)	(798)	(991)

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Exhibit 8.15 Voting in 2008 by Education and Income by Education

Voting by Education				
Education				
Voting	Grade School	High School Graduate	Some College	College Graduate
Voted	39.8%	69.7%	76.8%	87.2%
Did not vote	60.2%	30.3%	23.2%	12.8%
Total	100%	100%	100%	100%
(n)	(251)	(495)	(478)	(564)

Family Income by Education				
Education				
Family Income	Less Than High School	High School Graduate	Some College	College Graduate or Grad School
<$20,000	51.9%	27.3%	22.1%	7.6%
$20,000–$39,999	17.6%	27.1%	27.9%	13.7%
$40,000–$74,999	13.8%	27.1%	26.0%	24.6%
$75,000+	8.7%	18.5%	24.0%	54.1%
Total	100%	100%	100%	100%
(n)	(268)	(480)	(470)	(540)

Source: National Opinion Research Center (NORC). 2012. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Now try looking at the lower table, which is a bit more complex because it shows several levels of the dependent variable, family income. Try to see the effect that education has on income. Among the 268 grade school graduates surveyed (the first column on the left), you can see that 51.9%—more than half—have incomes less than $20,000 a year. Shifting to the high school graduates, the number in that lowest-income category has clearly fallen: The distribution has shifted some toward the higher income results. With some college, that trend continues, and for college graduates, you can see that 54.1% of them—more than half!—are making more than $75,000 a year. That's more than double (54.1 to 24.0) the percent of people who only did some college. Graduating from college pays off.

Encyclopedia Link

Read an overview of correlation.

So, education has a powerful effect on a person's chances for making a high income—which may be why many of you are reading this book right now!

When you read research reports and journal articles, you will find that social scientists usually judge the strength of association on the basis of more statistics than just a cross-tabulation table. A **measure of association** is a descriptive statistic used to summarize the strength of an association. One measure of association in cross-tabular analyses with ordinal variables is called **gamma**. The value of gamma ranges from −1 to +1. The closer a gamma value is to –1 or +1, the stronger the relationship between the two variables; a gamma of zero indicates that there is no relationship between the variables. Inferential statistics go further, addressing whether an association exists in the larger population from which the (random) sample was drawn. Even when the empirical association between two variables supports the researcher's hypothesis, it is possible that the association just resulted from the vagaries of random sampling. In a crosstab, estimation of this probability can be based on the inferential statistic, **chi-square**. The probability is customarily reported in a summary form such as $p < .05$, which can be translated as "The probability that the association resulted from chance is less than 5 out of 100 (5%)."

When the analyst feels reasonably confident (at least 95% confident, or $p < .05$) that an association did not result from chance, it is said that the association is statistically significant. **Statistical significance** basically means we conclude that the relationship is actually there; it's not a chance occurrence. Convention (and the desire to avoid concluding that an association exists in the population when it doesn't) dictates that the criterion be a probability of less than 5%. Statistical significance, though, doesn't equal substantive significance. That is, although the relationship is really occurring, not just happening accidentally, it may still not matter very much. It may be a minor part of what's happening.

Cross-tabulation (crosstab): In the simplest case, a bivariate (two-variable) distribution showing the distribution of one variable for each category of another variable; can also be elaborated using three or more variables.

Measure of association: A type of descriptive statistic that summarizes the strength of an

association.

Gamma: A measure of association that is sometimes used in cross-tabular analysis.

Chi-square: An inferential statistic used to test hypotheses about relationships between two or more variables in a cross-tabulation.

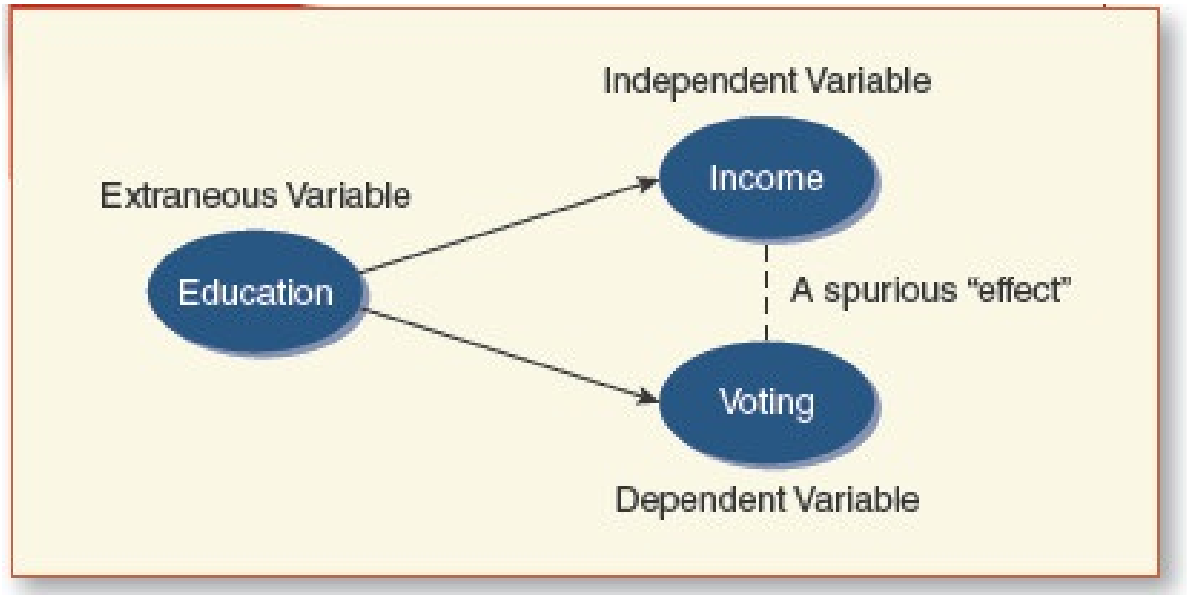
Statistical significance: The mathematical likelihood that an association is not the result of chance, judged by a criterion the analyst sets (often that the probability is less than 5 out of 100, or $p < .05$).

Controlling for a Third Variable

Cross-tabulation also can be used to study the relationship between three or more variables. The single most important reason for introducing a third variable is to see whether a bivariate relationship is spurious. A third, **extraneous variable**, for instance, may influence both the independent and dependent variables, creating an association between them that disappears when the extraneous variable is controlled. Ruling out possible extraneous variables helps strengthen considerably the conclusion that the relationship between the independent and dependent variables is causal—that it is nonspurious. In general, adding variables is termed ***elaboration analysis***: the process of introducing control or intervening variables into a bivariate relationship to better understand the relationship (Davis 1985; Rosenberg 1968).

For example, we have seen a positive association between incomes and the likelihood of voting; people with higher incomes are more likely to vote. But perhaps that association only exists because education influences both income and likelihood of voting; maybe when we control for education—that is, when we hold the value of education constant—we will find that there is no longer an association between income and voting. This possibility is represented by the hypothetical three-variable causal model in Exhibit 8.16, in which the arrows show that education influences both income and voting, thereby creating a relationship between the two. To test whether there is such an effect of education, we create the trivariate table in Exhibit 8.17, showing the bivariate crosstabs for various levels of education separately. This allows us to see if the income/voting relationship still exists after we hold education constant.

Exhibit 8.16 A Causal Model of a Spurious Effect



Source: National Opinion Research Center (NORC). 2006. General social survey. Chicago: National Opinion Research Center, University of Chicago.

The trivariate cross-tabulation in Exhibit 8.17 shows that the relationship between voting and income is *not* spurious because of the effect of education. The association between voting and income occurs in all three subtables. So our original hypothesis—that income as a social status indicator has an effect on voting—is not weakened.

Our goal in introducing you to cross-tabulation has been to help you think about the association among variables and to give you a relatively easy tool for describing association. To read most statistical reports and to conduct more sophisticated analyses of social data, you will have to extend your statistical knowledge, at least to include the technique of *regression* or *correlation analysis.* These statistics have many advantages over cross-tabulation—as well as some disadvantages. You will need to take a course in social statistics to become proficient in the use of statistics based on regression and correlation.

Extraneous variable: A variable that influences both the independent and dependent variables to create a spurious association between them that disappears when the extraneous variable is controlled.

Elaboration analysis: The process of introducing a third variable into an analysis to better understand—to elaborate—the bivariate (two-variable) relationship under consideration;

additional control variables also can be introduced.

Secondary Data Analysis

Secondary data analysis is the method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data. It has been an important social science methodology since the earliest days of social research, whether when Karl Marx (1967) reviewed government statistics in the Reading Room of the British Library during the 1850s to 1870s or Émile Durkheim (1966) analyzed official government cause-of-death data for his study of suicide rates throughout Europe in the late 19th century. With the advent of modern computers and, even more important, the Internet, secondary data analysis has become an increasingly accessible social research method. Literally thousands of large-scale data sets are now available for the secondary data analyst, often with no more effort than the few commands required to download the data set; a number of important data sets can even be analyzed directly on the web by users. The most common sources of **secondary data** are social science surveys and data collected by government agencies, often with survey research methods. It is also possible to reanalyze data that have been collected in experimental studies or with qualitative methods.

Exhibit 8.17 Voting in 2008 by Income and Education

	Family Income			
Voting	**<$20,000**	**$20,000–$39,999**	**$40,000–$74,999**	**$75,000+**
Education = < High school				
Voted	39.8%	38.1%	42.4%	33.3%
Did not vote	60.2%	61.9%	57.6%	66.7%
Total	100%	100%	100%	100%
(n)	(113)	(63)	(33)	(12)
Education = High school graduate				
Voted	57.1%	74.4%	73.0%	80.2%
Did not vote	42.9%	25.6%	27.0%	19.8%
Total	100%	100%	100%	100%
(n)	(119)	(121)	(122)	(81)
Education = Some college				
Voted	67.0%	75.8%	73.3%	90.4%
Did not vote	33.0%	24.2%	26.7%	9.6%
Total	100%	100%	100%	100%
(n)	(91)	(124)	(116)	(104)
Education = College graduate or graduate school				
Voted	71.8%	82.4%	85.8%	90.2%
Did not vote	28.2%	17.6%	14.2%	9.8%
Total	100%	100%	100%	100%
(n)	(39)	(74)	(127)	(275)

Source: National Opinion Research Center (NORC). 2006. General social survey. Chicago: National Opinion Research Center, University of Chicago.

Journal Link

Read about how researchers can use international secondary datasets to answer questions.

For several reasons, secondary analysis is very popular among professional social scientists. (1) Much of the groundwork involved in creating and testing measures with the data set has already been done. (2) Available data sets often

include many more measures and cases and reflect more rigorous research procedures than another researcher can afford to obtain. (3) Many social science projects collect data that can be used for questions that the primary researchers did not consider.

Sources range from data compiled by governmental units and private organizations for administrative purposes to data collected by social researchers that are then made available for reanalysis. Many important data sets are collected for the specific purpose of facilitating secondary data analysis. Government units from the U.S. Census Bureau to the U.S. Department of Housing and Urban Development; international organizations such as the United Nations, the Organisation for Economic Co-operation and Development (OECD), and the World Bank; and internationally involved organizations such as the Central Intelligence Agency (CIA) sponsor a substantial amount of social research that is intended for use by a broader community of social scientists. The National Opinion Research Corporation (NORC), with its General Social Survey (GSS), and the University of Michigan, with its Detroit Area Studies, are examples of academically based research efforts that gather data for social scientists to use in analyzing a range of social science research questions.

Since 1985, the GSS has participated in an International Social Survey Program that generates comparable data from 47 countries around the world (www.issp.org).

Many websites provide extensive collections of secondary data. Chief among these is the Inter-university Consortium for Political and Social Research (ICPSR) website at the University of Michigan. Searching for datasets at the ICPSR website can be as easy as entering in a search box the terms that describe your interests (see Exhibit 8.18).

Just one click at the ICPSR website will open the “Final Data” page offering a huge range of analyzable data sets. The ICPSR Academic consortium archives data sets online from major surveys and other social science research and makes them available for analysis by others.

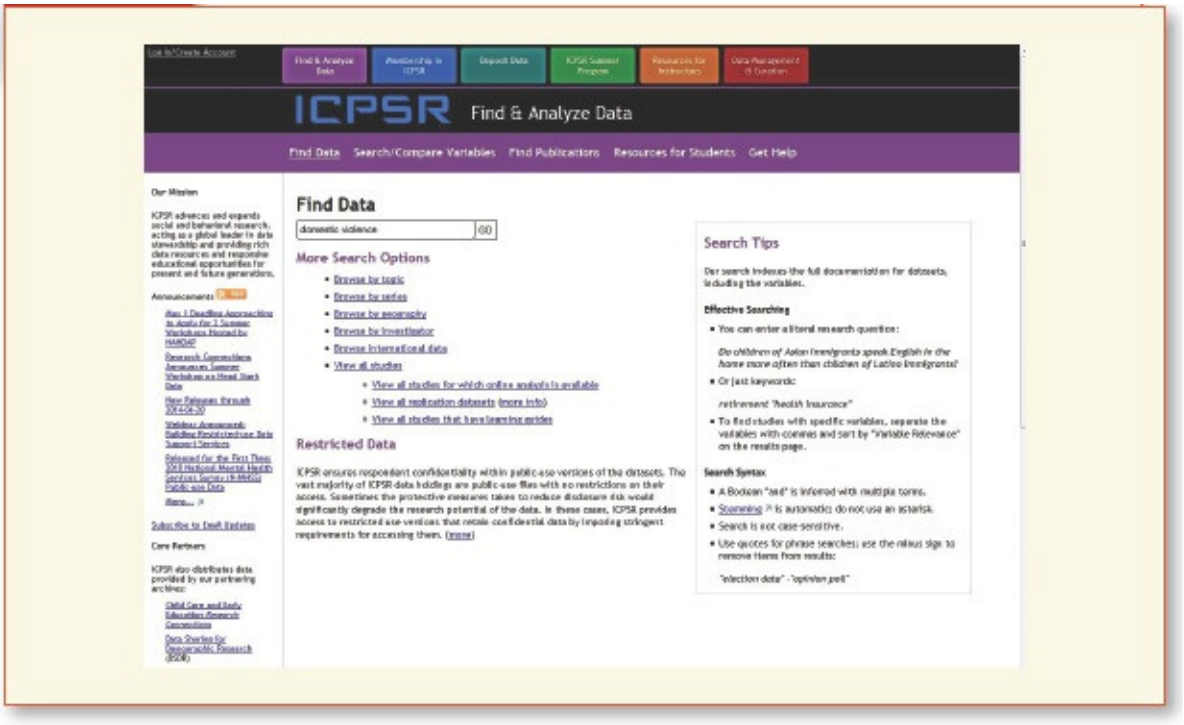
The University of California at Berkeley’s Survey Documentation and Analysis (SDA) archive provides several data sets from national omnibus surveys, as well as from U.S. Census microdata, from surveys on racial attitudes and prejudice, and from several labor and health surveys. The National Archive of Criminal

Justice Data is an excellent source of data in the area of criminal justice, although, like many other data collections, including key data from the U.S. Census, it is also available through the ICPSR. Much of the statistical data collected by U.S. federal government agencies can be accessed through the consolidated FedStats website, http://fedstats.sites.usa.gov.

The decennial population census by the U.S. Census Bureau is the single most important governmental data source, but many other data sets are collected by the U.S. Census and by other government agencies, including the U.S. Census Bureau's *Current Population Survey* and its Survey of Manufactures or the Bureau of Labor Statistics' Consumer Expenditure Survey. These government data sets typically are quantitative; in fact, the term *statistics*—state-istics—is derived from this type of data.

Details about some of the most important sources of secondary data will help you to think about the possibilities.

Exhibit 8.18 Search Screen: Domestic Violence



Source: Reprinted with permission from the Inter-University Consortium for Political and Social Research.

U.S. Census Bureau

The U.S. government has conducted a census of the population every 10 years since 1790; since 1940, this census has also included a census of housing. This decennial Census of Population and Housing is a rich source of social science data (Lavin 1994). The Census Bureau's monthly *Current Population Survey (CPS)* provides basic data on labor force activity that is then used in U.S. Bureau of Labor Statistics reports. The Census Bureau also collects data on agriculture, manufacturers, construction and other business, foreign countries, and foreign trade.

The U.S. Census of Population and Housing aims to survey one adult in every household in the United States. The basic *complete-count* census contains questions about household composition as well as ethnicity and income. Participation in the census is required by law, and confidentiality of the information obtained is mandated by law for 72 years after collection. Census data are reported for geographic units, including states, metropolitan areas, counties, census tracts (small, relatively permanent areas within counties), and even blocks. These different units allow units of analysis to be tailored to research questions.

Secondary data analysis The method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data.

Secondary data Previously collected data that are used in a new analysis.

Careers and Research

Claire Wulf Winiarek, MA, Director of Collaborative Policy Engagement

Source: Claire Wulf Winiarek

Claire Wulf Winiarek didn't set her sights on research methods as an undergraduate in political science and international relations at Baldwin College, nor as a masters student at Old Dominion University; her goal was to make a difference in public affairs. It still is. She is currently Director of Collaborative Policy Engagement at WellPoint, a Fortune 50 health insurance company based in Indianapolis, Indiana. Her previous positions include working for a Virginia member of the U.S. House of Representatives, coordinating grassroots international human rights advocacy for Amnesty International's North Africa Regional Action Network, and working as director of Public Policy and Research at Amerigroup's Office of Health Reform Integration.

Early in her career, Winiarek was surprised by the frequency with which she found herself

leveraging research methods. Whether she is analyzing draft legislation and proposed regulations, determining next year's department budget, or estimating potential growth while making the case for a new program, Winiarek has found that a strong foundation in research methods shapes her success. The increasing reliance of government and its private sector partners on data and evidence-based decision making continues to increase the importance of methodological expertise.

Policy work informed by research has made for a very rewarding career:

The potential for meaningful impact in the lives of everyday Americans is very real at the nexus of government and the private sector. Public policy, and how policy works in practice, has significant societal impact. I feel fortunate to help advance that nexus in a way that is informed not only by practice, evidence, and research, but also by the voice of those impacted.

Winiarek's advice for students seeking a career like hers is clear:

The information revolution is impacting all industries and sectors, as well as government and our communities. With this ever-growing and ever-richer set of information, today's professionals must have the know-how to understand and apply this data in a meaningful way. Research methods will create the critical and analytical foundation to meet the challenge, but internships or special research projects in your career field will inform that foundation with practical experience. Always look for that connection between research and reality

Bureau of Labor Statistics (BLS)

Another good source of data is the BLS of the U.S. Department of Labor, which collects and analyzes data on employment, earnings, prices, living conditions, industrial relations, productivity and technology, and occupational safety and health (U.S. Bureau of Labor Statistics 1991, 1997b). Some of these data are collected by the U.S. Census Bureau in the monthly *CPS*; other data are collected through surveys of establishments (U.S. Bureau of Labor Statistics 1997a).

The *CPS* provides a monthly employment and unemployment record for the United States, classified by age, sex, race, and other characteristics. The *CPS* uses a stratified random sample of about 60,000 households (with separate forms for about 120,000 individuals). Detailed questions are included to determine the precise labor force status (whether they are currently working or not) of each household member over the age of 16. Statistical reports are published each month in the BLS's *Monthly Labor Review* and can also be inspected at its website (http://stats.bls.gov). Data sets are available on computer tapes and disks from the BLS and services such as the ICPSR.

Inter-university Consortium for Political and Social Research

The University of Michigan’s ICPSR is the premier source of secondary data useful to social science researchers. ICPSR was founded in 1962 and now includes more than 640 colleges and universities and other institutions throughout the world. ICPSR archives the most extensive collection of social science data sets in the United States outside the federal government: More than 7,990 studies are represented in more than 500,000 files from 130 countries and from sources that range from U.S. government agencies such as the Census Bureau to international organizations such as the United Nations, social research organizations such as the National Opinion Research Center, and individual social scientists who have completed funded research projects.

Journal Link

Read an article that integrates multiple sources of secondary data.

In the United States, the ICPSR collection includes an expanding number of studies containing at least some qualitative data or measures coded from qualitative data (494 such studies by May 2011). Studies range from transcriptions of original handwritten and published materials relating to infant and child care from the beginning of the 20th century to World War II (LaRossa 1995) to transcripts of open-ended interviews with high school students involved in violent incidents (Lockwood 1996).

Human Relations Area Files

A unique source of qualitative data available for researchers in the United States is the Human Relations Area Files (HRAF) at Yale University. The HRAF has made anthropological reports available for international cross-cultural research since 1949 and currently contains more than 1,000,000 pages of information on more than 400 different cultural, ethnic, religious, and national groups (Ember & Ember 2011). If you are interested in cross-cultural research, it is well worth checking out the HRAF and exploring access options (reports can be accessed and searched online by those at affiliated institutions).

Secondary data analysis has some clear advantages (Rew et al. 2000: 226). It allows analyses of social processes in other inaccessible settings; it saves time and money; it allows the researcher to avoid data-collection problems; it facilitates comparison with other samples; it may allow inclusion of many more variables and a more diverse sample than otherwise would be feasible; it may allow data from multiple studies to be combined.

Conversely, with secondary data analysis, the researchers' cannot design data-collection methods that are best suited to answer his or her research question; he or she also cannot test and refine the methods to be used based on preliminary feedback from the population to be studied. Nor can the analyst engage in the iterative process of making observations, developing concepts, or making more observations and refining the concepts.

Secondary data analysis, then, inevitably involves a trade-off between the ease with which the research process can be initiated, and the specific hypotheses that can be tested. If the primary study was not designed adequately, the study may have to be abandoned (Riedel 2000: 53).

Research|Social Impact Link

Read more about the practical and ethical impacts of Big Data.

Data quality is always a concern with secondary data, even when the data are collected by an official government agency. Government actions result, at least in part, from political processes that may not have as their first priority the

design or maintenance of high-quality data for social scientific analysis.

Across national boundaries, different data-collection systems and definitions of key variables may have been used (Glover 1996). Census counts can be distorted by incorrect answers to census questions as well as by inadequate coverage of the entire population (Rives & Serow 1988: 32–35). For instance, national differences in the division of labor between genders within households can confuse the picture when comparing household earnings between nations without accounting for these differences (Jarvis 1997: 521).

Big Data

Big Data refers to data involving an entirely different order of magnitude than what we are used to thinking about as large data sets.

Big data analysis has only become possible with the development of very powerful information storage and very fast computing facilities.

For example (Mayer-Schönberger & Cukier 2013: 8–9): Facebook users upload more than 10 million photos every hour and leave a comment or click on a "like" button almost three billion times per day; YouTube users upload more than an hour of video every second; Twitter users were already sending more than 400 million tweets per day in 2012. If all this and other forms of stored information in the world were printed in books, one estimate in 2013 was that these books would cover the face of the Earth 52 layers thick. That's "Big."

Already, Big Data analyses are being used to predict the spread of flu, the price of airline tickets, and the behavior of consumers. Access to Big Data provides a new method for investigating the social world.

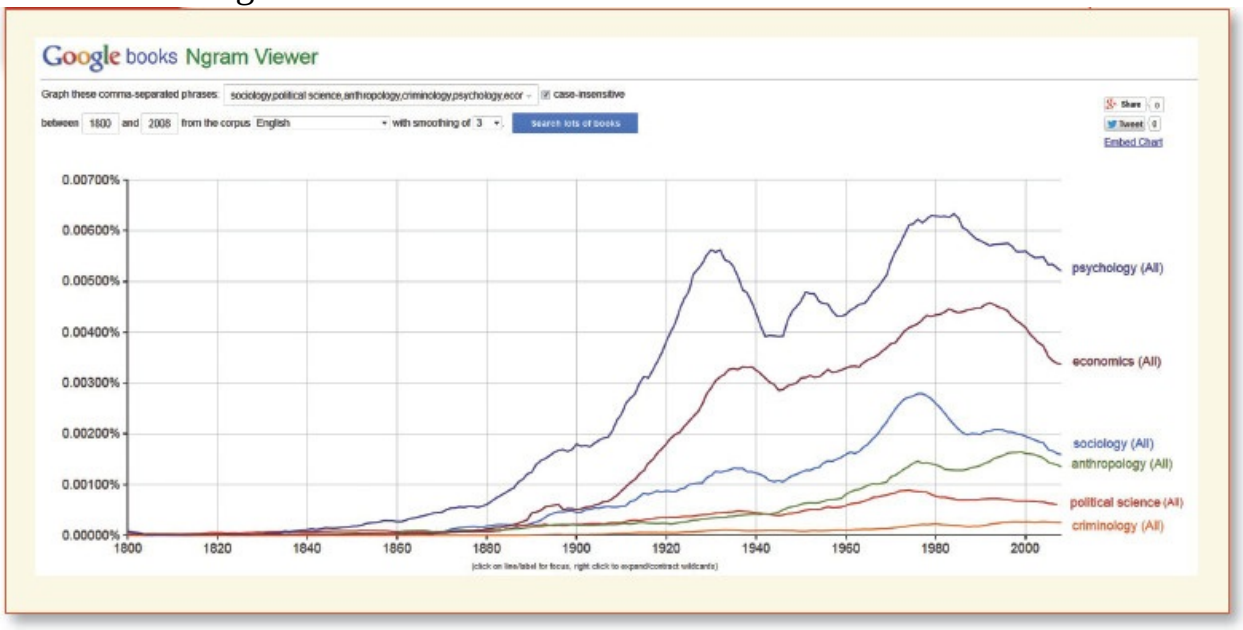
For instance, would you like to know how popular your discipline is? You can see how frequently the name of the discipline has appeared in all the books ever written in the world. It may surprise you to learn that it is possible right now to answer that question, although with two key limitations: we can only examine books written in English and in several other languages; and as of 2014 we are limited to "only" one quarter of all books ever published—a mere 30 million books (Aiden & Michel 2013: 16).

To check this out, go to the Google **Ngrams** site (https://books.google.com/ngrams), type in "sociology, political science, anthropology, criminology, psychology, economics," and check the "case-insensitive" box (and change the ending year to 2010). Exhibit 18.19 shows the resulting screen (if you don't obtain a graph, try using a different browser). Note that the height of a graph line represents the percentage that the term represents of all words in books published in each year, so a rising line means greater relative interest in the word, not simply more books being published. You can see that *psychology* emerges in the mid-19th century, whereas *sociology*, *economics*, *anthropology*, and *political science* appear in the latter part of that

century, and *criminology* arrives in the early 20th century. You can see that interest in sociology soared as the 1960s progressed, but then dropped off sharply in the 1980s. What else can you see in the graph?

The potential for Big Data is not just of academic interest. Jeremy Ginsberg and some colleagues (2009: 1012) at Google realized they could improve the response to the spread of flu around the world by taking advantage of the fact that about 90 million U.S. adults search online for information about specific illnesses each year. Ginsberg et al. started a collaboration with the U.S. Centers for Disease Control and Prevention (CDC), which collects data from about 2,700 health centers about patients' flu symptoms each year (Butler 2013: 155). By comparing this official CDC data with information from the Google searches, Ginsberg and his colleagues were able to develop a Big Data–based procedure for predicting the onset of the flu.

Exhibit 8.19 Ngram of Social Sciences



Source: Google Books. Ngram Viewer. http://books.google.com/ngrams.

But there were problems with the prediction. In the 2013 flu season, Google Flu Trends predicted a much higher peak level of flu than actually occurred. It seems to have been that widespread media coverage and the declaration of a public health emergency in New York led many more people than usual to search for

flu-related information, even though they were not experiencing symptoms themselves. Google has been refining its procedures to account for this problem, and other researchers have shifted their attention to analysis of flu-related "tweets" or to data from networks of thousands of volunteers who report symptoms experienced by family members to a central database (Butler 2013). So having incredible amounts of data does not solve all problems of sampling or measurement.

Sources of Big Data are increasing rapidly. One billion people use Facebook, thereby creating digital records that can, with appropriate arrangements, be analyzed to better understand social behavior (Aiden & Michel 2013: 12). Big Data are also generated by Global Positioning System (GPS) users, social media, smartphones, wristband health monitors, student postings, and even student activity in online education programs (Mayer-Schönberger & Cukier 2013: 90–96, 115). A new Big Data system records preemies' heart rate, respiration rate, temperature—what amounts to 1,260 data points per second—and can predict the onset of infection 24 hours before the appearance of overt symptoms (Mayer-Schönberger & Cukier 2013: 60). Public utilities, government agencies, and private companies can all learn about their customers from analyzing patterns revealed in their records.

The availability of Big Data makes possible the analysis of data from samples of a size previously unimaginable. Angela Bohn, Christian Buchta, Kurt Hornik, and Patrick Mair, in Austria and at Harvard in the United States, analyzed records on 438,851 Facebook users to explore the relation between friendship patterns and access to social capital (Bohn et al. 2014). Bohn et al. (2014: 32) started their analysis with data on 1,712 users—they didn't have the computer power to analyze more data—who were selected randomly over a 2-month study period, from about 1.3 million users who had agreed on Facebook to have their data used anonymously for such a study.

As you discovered when you started to check out the Google Ngrams site, having enormous sets of data readily available for analysis encourages exploration. "Rarely does [such a large amount of data] fit into neatly defined categories that are known at the outset. And the questions we want to ask often emerge only when we collect and work with the data we have" (Mayer-Schönberger & Cukier 2013: 45). Patterns discovered in Big Data may then suggest hypotheses that can be tested in causal experiments (Mayer-Schönberger & Cukier 2013: 65–66).

Big Data: Data produced or accessible in computer-readable form that is produced by people, available to social scientists, and manageable with today's computers.

Ngrams Frequency graphs produced by Google's database of all words printed in more than one third of the world's books over time (with coverage still expanding).

Ethical Issues in Statistical Analysis, Secondary Data Analysis, and Big Data

Using statistics ethically means, most importantly, being honest and open. Findings should be reported honestly, and the researcher should be open about the thinking that guided the decision to use particular statistics. Although this section has a mildly humorous title (after Darrell Huff's 1954 little classic, *How to Lie With Statistics)*, make no mistake about the intent: It is possible to distort social reality with statistics, and it is unethical to do so knowingly, even when the error results more from carelessness than to deceptive intent. There are a few basic rules to keep in mind:

Audio Link

Listen to information about ethical implications of the presentation of data.

- Inspect the shape of any distribution for which you report summary statistics to ensure that the statistics do not mislead your readers because of an unusual degree of skewness.
- When you create graphs, be sure to consider how the axes you choose may change the distribution's apparent shape; don't deceive your readers. You have already seen that it is possible to distort the shape of a distribution by manipulating the scale of axes, clustering categories inappropriately, and the like.
- Whenever you need to group data in a frequency distribution or graph, inspect the ungrouped distribution and then use a grouping procedure that does not distort the distribution's basic shape.
- Test hypotheses formulated in advance of data collection as they were originally stated. When evaluating associations between variables, it becomes very tempting to search around in the data until something interesting emerges. Social scientists sometimes call this a "fishing expedition." Although it's not wrong to examine data for unanticipated relationships, inevitably some relationships between variables will appear just on the basis of chance association alone. Exploratory analyses must be labeled in research reports as such.

- Be honest about the limitations of using survey data to test causal hypotheses. Finding that a hypothesized relationship is not altered by controlling for some other variables does not establish that the relationship is causal. There is always a possibility that some other variable that we did not think to control, or that was not even measured in the survey, has produced a spurious relationship between the independent and dependent variables in our hypothesis (Lieberson 1985). We have to think about the possibilities and be cautious in our causal conclusions.

Analysis of data collected by others, as well as content analysis of text, does not create the same potential for harm as does the collection of primary data, but neither ethical nor related political considerations can be ignored. First, because in most cases the secondary researchers did not collect the data, a key ethical obligation is to cite the original, principal investigators, as well as the data source, such as the ICPSR. Researchers who seek access to data sets available through the Council of European Social Science Data Archives (CESSDA) must often submit a request to the national data protection authority in the country (or countries) of interest (Johnson & Bullock 2009: 214).

Subject confidentiality is a key concern when original records are analyzed. Whenever possible, all information that could identify individuals should be removed from the records to be analyzed so that no link is possible to the identities of living subjects or the living descendants of subjects (Huston & Naylor 1996: 1698). When you use data that have already been archived, you need to find out what procedures were used to preserve subject confidentiality. The work required to ensure subject confidentiality probably will have been done for you by the data archivist. For example, the ICPSR examines carefully all data deposited in the archive for the possibility of disclosure risk. All data that might be used to identify respondents are altered to ensure confidentiality, including removal of information such as birth dates or service dates, specific incomes, or place of residence that could be used to identify subjects indirectly (see http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html). If all information that could be used in any way to identify respondents cannot be removed from a data set without diminishing data set quality (e.g., by preventing links to other essential data records), ICPSR restricts access to the data and requires that investigators agree to conditions of use that preserve subject confidentiality. Those who violate confidentiality may be subject to a scientific misconduct investigation by their home institution at the request of

ICPSR (Johnson & Bullock 2009: 218). The UK Data Archive provides more information about confidentiality and other human subjects protection issues at www.dta-archive.ac.uk/create-manage/consent-ethics.

Conclusion

With some simple statistics (means, standard deviations, and the like), a researcher can describe social phenomena, identify relationships among them, explore the reasons for these relationships (especially through elaboration), and test hypotheses about them. Statistics—carefully constructed numbers that describe an entire population of data—are amazingly helpful in giving a simple summation of complex situations. Statistics provide a remarkably useful tool for developing our understanding of the social world, a tool that we can use both to test our ideas and to generate new ones.

Unfortunately, to the uninitiated, the use of statistics can seem to end debate right there—one can't argue with the numbers. But you now know better. Numbers are worthless if the methods used to generate the data are not valid, and numbers can be misleading if they are not used appropriately, considering the type of data to which they are applied. In a very poor town with one wealthy family, the mean income may be fairly high—but grossly misleading. And even assuming valid methods and proper use of statistics, there's one more critical step, because the numbers do not speak for themselves. Ultimately, how we interpret and report statistics determines their usefulness.

Key Terms

Bar chart 171
Base number (*N*) 174
Big data 190
Bimodal 177
Central tendency 169
Chi-square 185
Cross-tabulation (crosstab) 182
Data cleaning 168
Descriptive statistics 168
Elaboration analysis 185
Extraneous variable 185
Frequency distribution 173
Frequency polygon 172
Gamma 184
Histogram 171
Inferential statistics 168
Interquartile range 180
Mean 178
Measure of association 184
Median 177
Mode (probability average) 177
Ngrams 191
Normal distribution 181
Outlier 180
Percentage 173
Quantitative data analysis 167
Quartiles 180
Range 180
Secondary data 187
Secondary data analysis 186
Skewness 169
Standard deviation 181
Statistic 168
Statistical significance 185
Unimodal 177

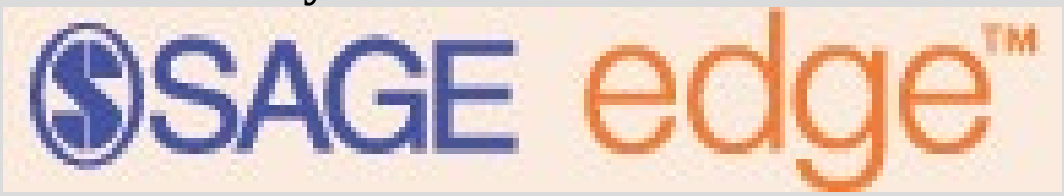
Variability 169
Variance 180

Highlights

- Data entry options include direct collection of data through a computer, use of scannable data entry forms, and use of data-entry software. All data should be cleaned during the data-entry process.
- Use of secondary data can save considerable time and resources but may limit data analysis possibilities.
- Bar charts, histograms, and frequency polygons are useful for describing the shape of distributions. Care must be taken with graphic displays to avoid distorting a distribution's apparent shape.
- Frequency distributions display variation in a form that can be easily inspected and described. Values should be grouped in frequency distributions in a way that does not alter the shape of the distribution. Following several guidelines can reduce the risk of problems.
- Summary statistics are often used to describe the central tendency and variability of distributions. The appropriateness of the mode, mean, and median vary with a variable's level of measurement, the distribution's shape, and the purpose of the summary.
- The variance and standard deviation summarize variability around the mean. The interquartile range is usually preferable to the range to indicate the interval spanned by cases because the effect of outliers on the range. The degree of skewness of a distribution is usually described in words rather than with a summary statistic.
- Cell frequencies in cross-tabulation should normally be converted to percentages within the categories of the independent variable. A cross-tabulation can be used to determine the existence, strength, direction, and pattern of an association.
- Elaboration analysis can be used in cross-tabular analysis to test for spurious relationships.
- Inferential statistics are used with sample-based data to estimate the confidence that can be placed in a statistical estimate of a population parameter. Estimates of the probability that an association between variables may have occurred on the basis of chance are also based on inferential statistics.
- Secondary data analysis enables researchers to use existing data to investigate new research questions and can be obtained easily from many sources.

- Big Data analysis involves the statistical analysis of patterns in extremely large datasets generated by records of social activity.
- Honesty and openness are the key ethical principles that should guide data summaries.

Student Study Site



The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. We presented in this chapter several examples of bivariate and trivariate cross-tabulations involving voting in the 2008 presidential election. What additional influences would you recommend examining to explain voting in elections? Suggest some additional independent variables for bivariate analyses with voting, as well as several additional control variables to be used in three-variable crosstabs.
2. When should we control just to be honest? Should social researchers be expected to investigate alternative explanations for their findings? Should they be expected to check to see if the associations they find occur for different subgroups in their samples? Justify your answers.

Finding Research

1. Do a web search for information on a social science subject in which you are interested. How much of the information you find relies on statistics as a tool for understanding the subject? How do statistics allow researchers to test their ideas about the subject and generate new ideas? Write your findings in a brief report, referring to the websites on which you relied.
2. The National Bureau of Economic Research provides many graphs and numeric tables about current economic conditions (www.nber.org/). Review some of these presentations. Which displays are most effective in conveying information? Summarize what you can learn from this site about economic conditions.

Critiquing Research

1. Become a media critic. For the next week, scan a newspaper or some magazines for statistics. How many articles can you find that use frequency distributions, graphs, and the summary statistics introduced in this chapter? Are these statistics used appropriately and interpreted correctly? Would any other statistics have been preferable or useful in addition to those presented?

Doing Research

Exhibit 8.20 Is Child Care Important? By Gender and Marital Status

	MEN		WOMEN	
	Single	Married	Single	Married
Not important	54%	48%	33%	12%
Somewhat important	24%	30%	45%	31%
Very important	22%	22%	22%	57%
	100%	100%	100%	100%
n =	(125)	(218)	(51)	(161)

Source: Created by Daniel F. Chambliss for this volume.

1. Create frequency distributions from lists in U.S. Census Bureau reports on the characteristics of states, cities, or counties or any similar listing of data for at least 100 cases (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml). You will have to decide on a grouping scheme for the distribution of variables, such as average age and population size; how to deal with outliers in the frequency distribution; and how to categorize qualitative variables, such as the predominant occupation. Decide what summary statistics to use for each variable. How well were the features of each distribution represented by the summary statistics? Describe the shape of each distribution. Propose a hypothesis involving two of these variables, and develop a crosstab to evaluate the support for this hypothesis. Describe each relationship in terms of the four aspects of an association after converting cell frequencies to percentages in each table within the categories of the independent variable. Does the hypothesis appear to have been supported?

2. Exhibit 8.20 is a three-variable table created with survey data from 355 employees hired during the previous year at a large telecommunications company. Employees were asked if the presence of on-site child care at the company's offices was important in their decision to join the company.

   Reading the table:
   1. Does gender affect attitudes?
   2. Does marital status affect attitudes?
   3. Which of the preceding two variables matters more?
   4. Does being married affect men's attitudes more than women's?

3. If you have access to the SPSS statistical program, you can analyze data contained in the 2012 General Social Survey (GSS) file on the Study Site for this text.

   Develop a description of the basic social and demographic characteristics of the U.S. population in 2012. Examine each characteristic with three statistical techniques: a graph, a

frequency distribution, and a measure of central tendency (and a measure of variation, if appropriate).

    1. From the menu, select “Graphs” and then “Legacy Dialogs and Bar.” Select “Simple Define” [Marital—Category Axis]. Bars represent % of cases. Select “Options” (do not display groups defined by missing values). Finally, select “Histogram” for each of the variables [EDUC, EARNRS, TVHOURS, ATTEND].
    2. Describe the distribution of each variable.
    3. Generate frequency distributions and descriptive statistics for these variables. From the menu, select “Analyze/ Descriptive Statistics/Frequencies.” From the “Frequencies” window, set MARITAL, EDUC, EARNRS, TVHOURS, ATTEND. For the “Statistics,” choose the mean, median, range, and standard deviation.
    4. Which statistics are appropriate to summarize the central tendency and variation of each variable? Do the values of any of these statistics surprise you?

4. Try describing relationships with support for capital punishment by using graphs. Select two relationships you identified in previous exercises and represent them in graphic form. Try drawing the graphs on lined paper (graph paper is preferable).

Ethics Questions

1. Review the frequency distributions and graphs in this chapter. Change one of these data displays so that you are "lying with statistics." (You might consider using the graphic technique discussed by Orcutt & Turner 1993.)
2. Consider the relationship between voting and income that is presented in Exhibit 8.13. What third variables do you think should be controlled in the analysis to understand better the basis for this relationship? How might social policies be affected by finding out that this relationship was caused by differences in neighborhood of residence rather than by income itself?

Video Interview Questions

Listen to the interview with Peter Marsdan for Chapter 8 at edge.sagepub.com/chamblissmssw5e.

1. What are the three goals of the General Social Survey (GSS)?
2. When was the first GSS conducted? Who developed the GSS concept?

Chapter 9 Qualitative Methods Observing, Participating, Listening

©iStockphoto.com/FilippoBacci

Learning Objectives

1. Identify the circumstances that make qualitative methods most useful.
2. Describe the features of qualitative research that most distinguish it from quantitative research.
3. Define the methods of ethnography and netnography.
4. Compare the advantages and disadvantages of each participant observer role.
5. Discuss the major challenges at each stage of a field research project.
6. Explain how to record and analyze field notes.
7. Describe the process of intensive interviewing, and compare it to the process of interviewing in survey research.
8. Discuss the advantages of focus group research, and identify particular challenges focus group researchers face.
9. Identify the major ethical challenges faced by qualitative researchers, and discuss one qualitative research project that posed particular ethical concerns.

Qualitative research goes straight to where people live—and die:

We see what those poor bastards go through. Seriously, when [a dying medical patient has] been resuscitated nine or ten times and their chest looks like raw meat, they've been fried from being defibrillated, they've had their chest pumped on, they've got a flat chest because their ribs are no more connected to their sternum. . . . You know this guy doesn't have a chance in hell. I mean, he's already blown out, squash, herniated his brain, he doesn't have any spontaneous respirations, he's flat EEGs. You take care of him for eight hours, you know that this person is not viable, and you feel for him and you feel for the family. . . . When you're resuscitating somebody and they get no response going into the code for an hour, and now has no EKG, no heart tracing, pupils are blown, fixed, no spontaneous respiration, blood gases are out in the ozone. . . . you are the one that's going to turn to the resident and say, "Don't you think this is about it, don't you think we should call this?" (interview, as cited in Chambliss 1996: 164)

Throughout this chapter, you will learn that some of our greatest insights into social processes can result from what appear to be very ordinary activities: observing, participating, listening, and talking. But you will also learn that qualitative research is much more than just doing what comes naturally: Qualitative researchers must observe keenly, take notes systematically, question respondents strategically, and prepare to spend more time and invest more of

their whole selves than often occurs with experiments or surveys.

We begin with an overview of the major features of qualitative research. The next section discusses participant observation research, which is the most distinctive qualitative method. We then discuss intensive interviewing—a type of interviewing that qualifies as qualitative rather than quantitative research—and focus groups, an increasingly popular qualitative method. The last two sections discuss how to analyze qualitative data and make ethical decisions in qualitative research.

What Are Qualitative Methods?

Qualitative methods refer to several distinct research activities: participant observation, intensive interviewing, and focus groups.

Although these three qualitative designs differ in many respects, they share several features, in addition to the collection of qualitative data itself, that distinguish them from experimental and survey research designs (Denzin & Lincoln 1994; Maxwell 1996; Wolcott 1995):

- Qualitative researchers typically begin with an exploratory research question about what people think and how they act, and why, in some social setting. This research approach is primarily inductive.
- The designs *focus on previously unstudied processes and unanticipated phenomena* because previously unstudied attitudes and actions can't adequately be understood with a structured set of questions or within a highly controlled experiment.
- Qualitative designs have an *orientation to social context*, to the interconnections between social phenomena rather than to their discrete features.
- The designs *focus on human subjectivity*, on the meanings that participants attach to events and that people give to their lives.
- The designs have a *sensitivity to the subjective* role of the researcher. Qualitative researchers consider themselves as necessarily part of the social process being studied and, therefore, keep track of their own actions in, and reactions to, that social process.

Case Study: Beyond Caring

In preparing to write his 1996 book *Beyond Caring: Hospitals, Nurses, and the Social Organization of Ethics,* Dan Chambliss spent many months, spread over 12 years, studying hospital nurses at work. Observing in several different hospitals, in different regions of the United States, Chambliss watched countless operations and emergency room crises, but he also sat up nights chatting with nurses on geriatric floors (specializing in the care of old people) and quietly watched for hours at a time while nurses did postoperative care, bathed patients, helped patients walk down the hall, or just met with each other and with doctors, technicians, and aides to discuss the day's work. He also conducted more than 100 formal interviews, averaging 1.5 hours or more each; he attended birthday parties and softball games and saw nurses in social situations as well as at professional conferences. This project exemplifies **field research**, which combines various forms of qualitative research.

Research|Social Impact Link

Read about how rich qualitative data contrasts with quantitative statistics.

The resulting data are nothing like the clean list of responses given to a survey questionnaire. Instead, Chambliss (1996) wrote his book from boxes full of notes on his observations, such as these:

> [Today I witnessed] the needle injection of local anesthetic into a newborn (3 weeks) baby's skull, so they could remove a shunt. The two residents doing it discussed whether a local anesthetic would be sufficient; a general [anesthetic] would be dangerous. One said, "I can do it if you can." This exchange was carried out a couple of times. A nurse (man) stroked the infant's hand, talked softly to it, and calmed it immediately as they were setting up, putting in the IVs—hard to do, the veins are so small.
>
> The resident injected the local anesthetic. Everyone around was affected by the immediate widening of the baby's eyes as the needle first went in, and then the screaming. The resident doing it, though, was absolutely concentrated on the task. At one point the female resident mentioned her

> concern, saying something about the whole point of anesthetic is to lessen pain, not to increase it. The baby was put in pain, couldn't have known any reason for it, was helpless to resist. [Field Notes] (pp. 135–136)

So fieldwork involves, at its simplest, spending time with people in their own settings, watching them do what they do. Gary Allen Fine, a prominent field researcher, has studied Little League baseball, restaurant kitchens, high school debate teams, and people who hunt for mushrooms, to name a few settings. Chambliss had complete access to the working (and sometimes personal) lives of the nurses he studied.

Such research obviously requires a huge investment of time. Chambliss moved his residence several times during his research, living in apartments near the medical centers that he studied. He built his entire schedule, for months on end, around the opportunities for seeing often unseen things—emergency resuscitations, hidden malpractice, even the boredom of some nursing work.

Journal Link

Read how field research was used to examine social behavior after Hurricane Katrina.

But the investment can be worth the cost. Chambliss's (1996) early research on nurses primarily relied on tape-recorded interviews:

> These [interviews] produced many dramatic stories and often confirmed theories I already held, but as I began to spend more time in hospitals I began to doubt the veracity of interviews. I began to see how the interviews were a reflection of my interests as much as of my subjects' lives. The stories told were more exciting than the ordinary drudgery I saw; the nurses described in stories seemed more committed and courageous than some of those I actually watched. Interviewees told what they noticed and remembered, which I discovered to be a highly selective version of what actually occurred. Much of life, I found, consists precisely in not noticing what one does all the time. "There aren't any ethical problems here I can think of," said a pediatric research nurse mentioned earlier; "You should

> talk with people on the ethics committee," said nurses gathered outside the room of an AIDS patient. (pp. 194)

Chambliss wanted to learn about nurses, so in a sense he did the obvious: He worked and talked with nurses, many of them, over a long period. But he also took care to study a variety of hospitals and different services within hospitals; he also "sampled" different times of the day and night and different kinds of patients. True, such research is inductive, and the researcher is open to surprises; Chambliss couldn't run controlled experiments or easily isolate independent and dependent variables. But even the most unstructured kind of research still adheres to the basic discipline of scientific method.

There are many different qualitative methods. Here we first focus attention on three qualitative methods that illustrate the flexibility of this approach: ethnography, netnography, and ethnomethodology. We then discuss how to collect data using three different qualitative strategies: participant observation, intensive interviewing, and focus groups. In Chapter 10, you will learn how researchers analyze data collected with these methods.

Qualitative methods: Methods, such as participant observation, intensive interviewing, and focus groups, that are designed to capture social life as participants experience it rather than in categories the researcher predetermines. These methods typically involve exploratory research questions, inductive reasoning, an orientation to social context, and a focus on human subjectivity and the meanings participants attach to events and to their lives.

Field research: Research in which natural social processes are studied as they happen and left relatively undisturbed.

Ethnography

Field research borrows heavily from a long-standing traditional method of anthropological studies called **ethnography**. Ethnography is the study of a culture or cultures that some group of people share (Van Maanen 1995: 4). As a method, it usually refers to participant observation by a single investigator immersed in the group for a long time (often 1 or more years). Ethnographic research can also be termed *naturalistic* because it seeks to describe and understand the natural social world as it really is, in all its richness and detail. Anthropological field research has traditionally been ethnographic, and much sociological fieldwork shares these same characteristics. But there are no particular methodological techniques associated with ethnography other than just "being there." The analytic process relies on the thoroughness and insight of the researcher to "tell us like it is" in the setting, as she or he experienced it.

Video Link

Watch a clip about ethnography.

Code of the Street, Elijah Anderson's (2000: 11) award-winning study of Philadelphia's inner city, captures the flavor of this approach:

> My primary aim in this work is to render ethnographically the social and cultural dynamics of the interpersonal violence that is currently undermining the quality of life of too many urban neighborhoods. . . . How do the people of the setting perceive their situation? What assumptions do they bring to their decision making?

Journal Link

Read about an ethnographic study exploring community identity and the 9/11 attacks.

Anderson's methods are described in the book's preface: participant observation, including direct observation and in-depth interviews; impressionistic materials drawn from various social settings around the city; and interviews with a wide variety of people. Like most traditional ethnographers, Anderson (2000) describes his concern with being "as objective as possible" and using his training, as other ethnographers do, "to look for and to recognize underlying assumptions, their own and those of their subjects, and to try to override the former and uncover the latter" (p. 11).

From analysis of the data obtained in these ways, a rich description emerges of life in the inner city. Although we often do not "hear" the residents speak, we feel the community's pain in Anderson's (2000) description of "the aftermath of death":

> When a young life is cut down, almost everyone goes into mourning. The first thing that happens is that a crowd gathers about the site of the shooting or the incident. The police then arrive, drawing more of a crowd. Since such a death often occurs close to the victim's house, his mother or his close relatives and friends may be on the scene of the killing. When they arrive, the women and girls often wail and moan, crying out their grief for all to hear, while the young men simply look on, in studied silence. . . . Soon the ambulance arrives. (p. 138)

Anderson (2000) uses these descriptions as a foundation on which he develops the key concepts in his analysis, such as "code of the street":

> The "code of the street" is not the goal or product of any individual's actions but is the fabric of everyday life, a vivid and pressing milieu within which all local residents must shape their personal routines, income strategies, and orientations to schooling, as well as their mating, parenting, and neighbor relations. (p. 326)

Anderson's (2003) report on his Jelly's Bar study illustrates how an ethnographic analysis deepened as he became more socially integrated into the Jelly's Bar group. He thus became more successful at "blending the local knowledge one has learned with what we already know sociologically about such settings" (p.

236).

> I engaged the denizens of the corner and wrote detailed field notes about my experiences, and from time to time looked for patterns and relationships in my notes. In this way, an understanding of the setting came to me in time, especially as I participated more fully in the life of the corner and wrote my field notes about my experiences; as my notes accumulated, and as I reviewed them occasionally and supplemented them with conceptual memos to myself, their meanings became more clear, while even more questions emerged. (p. 224)

Recently such ethnographic work has been flourishing, with a host of talented young researchers doing fascinating studies: Matt Desmond's participant observations of wildland firefighters and the "country masculinity" they embody (Desmond 2007); Alice Goffman's heartrending descriptions of young black men constantly "on the run" from an all-surveilling criminal justice system (Goffman 2014); Colin Jerolmack's phenomenology of pigeon breeders in New York and Berlin (Jerolmack 2007, 2009); Claudio Benzecry's witty evocation of the lives and passions of Argentine opera fanatics (Benzecry 2011)—all show that even in this age of so much computer-driven research, the tradition, born from anthropology and sociology, of close-up qualitative fieldwork is anything but dead.

Ethnography: The study and systematic recording of human cultures.

Careers and Research

Amanda Aykanian, Research Associate, Advocates for Human Potential

Source: Amanda Aykanian

Amanda Aykanian majored in psychology at Framingham State University and found that she enjoyed the routine and organization of research. She wrote an undergrad thesis to answer the research question: How does the way in which course content is presented affect students' feelings about the content and the rate at which they retain it?

After graduating, Aykanian didn't want to go to graduate school right away; instead she wanted to explore her interests and get a sense of what she could do with research. Advocates for Human Potential (AHP) was the last research assistant (RA) job that Aykanian applied for. Her initial tasks as an RA at AHP ranged from taking notes, writing agendas, and assembling project materials to entering research data, cleaning data, and proofing reports. As she contributed more to project reports, she began to think about data from a more theoretical standpoint.

During 7 years at AHP, Aykanian has helped lead program evaluation research, design surveys and write survey questions, conduct phone and qualitative interviews, and lead focus groups. Her program evaluation research almost always uses a mixed-methods approach, so Aykanian has learned a lot about how qualitative and quantitative methods can complement each other. She has received a lot of on-the-job training in data analysis and has learned how to think about and write a proposal in response to federal funding opportunities.

Aykanian was promoted to research associate and describes her current role as part program evaluation coordinator and part data analyst. She has also returned to graduate school, earning a master's degree in applied sociology and then starting a PhD program in social welfare.

Netnography

As you know from social media like Facebook, *communities* now refer not only to people in a common physical location, but also to relationships that develop online. Online communities may be formed by persons with similar interests or backgrounds, perhaps to create new social relationships that location or schedules did not permit, or to supplement relationships that emerge in a course of work or school or other ongoing social activities. Like communities of people who interact face-to-face, online communities can develop a culture and become sources of identification and attachment (Kozinets 2010: 14–15). And like physical communities, researchers can study online communities through immersion in the group for an extended period. **Netnography**, also termed ***cyberethnography* and *virtual ethnography*** (James & Busher 2009: 34–35), is the use of ethnographic methods to study online communities.

In some respects, netnography is similar to traditional ethnography. The researcher prepares to enter the field by becoming familiar with online communities and their language and customs, formulating an exploratory research question about social processes or orientation in that setting, selecting an appropriate community to study. Unlike in-person ethnographies, netnographies can focus on communities whose members are physically distant and dispersed. The selected community should be relevant to the research question, involve frequent communication among actively engaged members, and have a number of participants who, as a result, generate a rich body of textual data (Kozinets 2010: 89).

The netnographer's self-introduction should be clear and friendly. Robert Kozinets (2010) provides the following example written about the online discussion space alt.coffee:

> I've been lurking here for a while, studying online coffee culture on alt.coffee, learning a lot, and enjoying it very much. . . . I just wanted to pop out of lurker status to let you know I am here. . . . I will be wanting to quote some of the great posts that have appeared here, and I will contact the individuals by personal e-mail who posted them to ask their permission to quote them. I also will be making the document on coffee culture available to any interested members of the newsgroup for their perusal and comments

—to make sure I get things right. (p. 93)

A netnographer must keep both observational and reflective field notes but, unlike a traditional ethnographer, can return to review the original data—the posted text—long after it was produced. The data can then be coded, annotated with the researcher's interpretations, checked against new data to evaluate the persistence of social patterns, and used to develop a theory that is grounded in the data.

Netnography (cyberethnography and virtual ethnography): The use of ethnographic methods to study online communities.

Research in the News

Can Taping Interviews Capture A Trend?

Sociologist Eric Klinenberg used qualitative interviewing to debunk assumptions about individuals who live alone. Klinenberg interviewed 300 people living alone during a 10-year period. What did he find? People who live alone are more social and less isolated. Intensive interviewing revealed older individuals expressing a desire for independence and single living. Economics greatly affect the ability to live alone, and cultures all over the globe are seeing an increase in solo living.

1. Why might have Klinenberg used qualitative interviewing in this research rather than a quantitative survey? Explain why qualitative interviewing may have been more suited to identifying this misconception.
2. How could you design a study to explore this issue in different cultures? What interpretations could other cultures have of living alone?

News Source: Klinenberg, Eric. 2012. One's a crowd. *New York Times*, February 5: SR4.

Ethnomethodology

Ethnomethodology, a notable variation of fieldwork, studies the way that participants construct the social world in which they live—how they "create reality"—rather than trying to describe the social world objectively. In fact, ethnomethodologists do not necessarily believe that we can find an objective reality; instead, how participants come to create and sustain a sense of "reality" is the focus of study. In the words of Jaber F. Gubrium and James A. Holstein (1997), in ethnomethodology, compared to the naturalistic orientation of ethnography,

> the focus shifts from the scenic features of everyday life onto the ways through which the world comes to be experienced as real, concrete, factual, and "out there." An interest in members' methods of constituting their worlds supersedes the naturalistic project of describing members' worlds as they know them. (p. 41)

Unlike the ethnographic analyst, who seeks to describe the social world as the participants see it, the ethnomethodological analyst seeks to maintain some distance from that world. The ethnomethodologist views a "code" of conduct, like that described by Anderson (2003), not as a description of a real normative force that constrains social action but as the way that people in the setting create a sense of order and social structure (Gubrium & Holstein 1997: 44–45). The ethnomethodologist focuses on how reality is constructed, not on what it is.

Ethnomethodology: A qualitative research method focused on the way that participants in a social setting create and sustain a sense of reality.

How Does Participant Observation Become a Research Method?

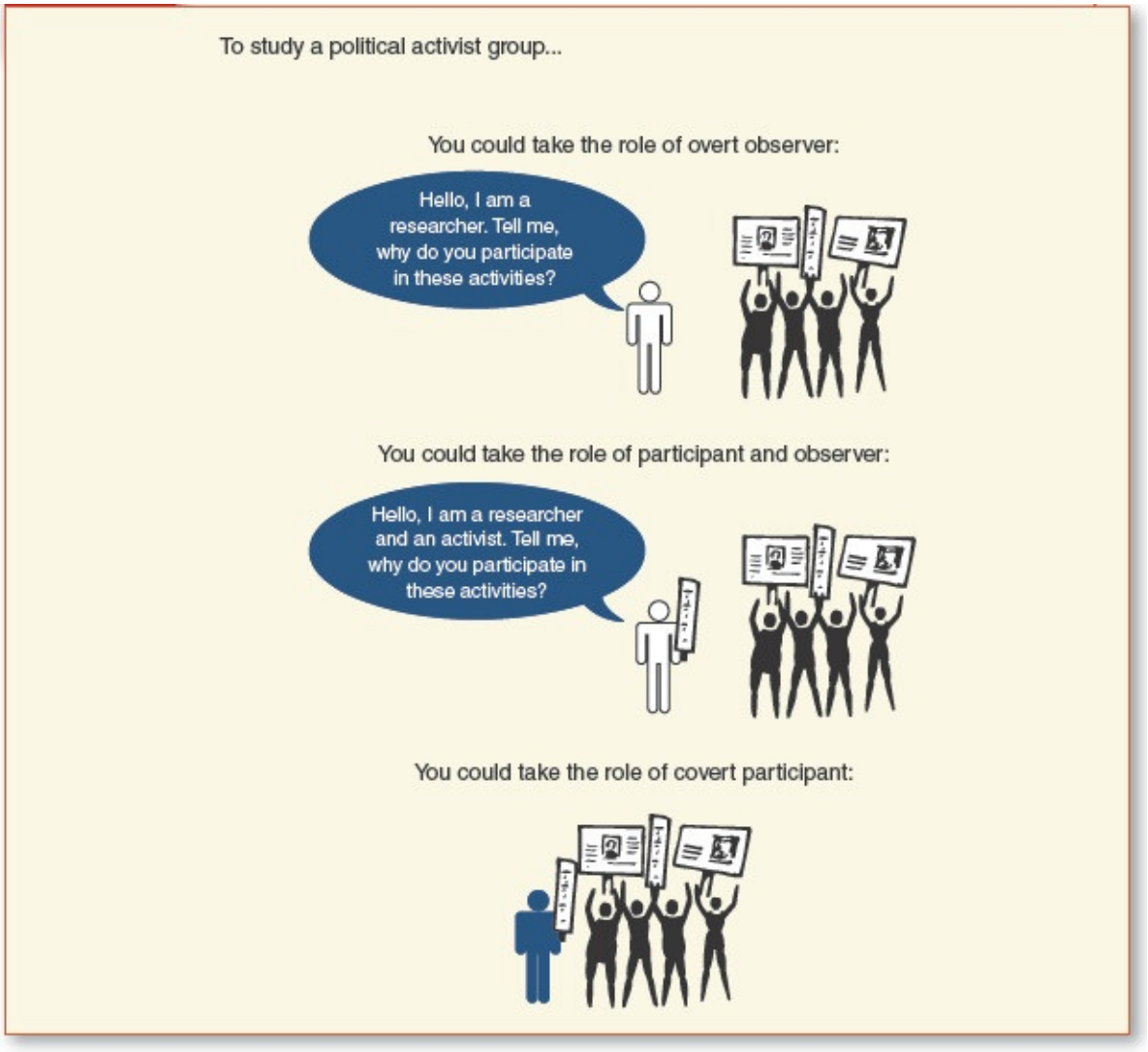
Desmond used **participant observation**, working as a firefighter, to study the teams of "hotshots" who fight grass and forest fires because it would leave natural social processes, in their natural setting, relatively undisturbed. Such fieldwork or field research, going out to where people really live and work, is a means for seeing the social world as the research subjects see it, in its totality, and for understanding subjects' interpretations of that world (Wolcott 1995: 66). Participant observers seek to avoid the artificiality of experimental designs and the unnatural structured questioning of survey research (Koegel 1987: 8). This method encourages consideration of the context in which social interaction occurs, of the complex and interconnected nature of social relations, and of the sequencing of events (Bogdewic 1999: 49). Through it, we can understand the *mechanisms* (one of the criteria for establishing cause) of social life.

In his study of nursing homes, Timothy Diamond (1992) explained how his exploratory research question led him to adopt the method of participant observation:

> How does the work of caretaking become defined and get reproduced day in and day out as a business? . . . The everyday world of Ina and Aileen and their co-workers, and that of the people they tend. . . . I wanted to collect stories and to experience situations like those Ina and Aileen had begun to describe. I decided that . . . I would go inside to experience the work myself. (p. 5)

The term *participant observer* actually represents a continuum of roles (Exhibit 9.1), ranging from being a complete observer who does not participate in group activities and is publicly defined as a researcher to being a covert participant who acts just like other group members and does not disclose his or her research role. Many field researchers develop a role between these extremes, publicly acknowledging being a researcher but nonetheless participating in group activities.

Exhibit 9.1 The Observational Continuum


To study a political activist group...
You could take the role of overt observer:
Hello, I am a researcher. Tell me, why do you participate in these activities?
You could take the role of participant and observer:
Hello, I am a researcher and an activist. Tell me, why do you participate in these activities?
You could take the role of covert participant:

Choosing a Role

The first concern of all participant observers is deciding what balance to strike between observing and participating and whether to reveal their roles as researchers. These decisions must consider the specifics of the social situation being studied, the researcher's own background and personality, the larger sociopolitical context, and ethical concerns. Which balance of participating and observing is most appropriate also changes during most projects—often many times.

Complete Observation

In **complete observation**, researchers try to see things as they happen, without actively participating in these events. Chambliss watched nurses closely, but he never bathed a patient, changed a dressing, started an intravenous line, or told a family that their loved one had died. Once during an emergency surgery for a ruptured ectopic pregnancy—a drastic, immediately life-threatening event—a surgeon ordered him to "put in a Foley" (a urinary catheter), but a nurse quickly said, "He's a researcher, I'll do it." Of course, at the same time as observing a setting, researchers must consider the ways in which their presence as observers itself alters the social situation being observed. Such **reactive effects** occur because it is not "natural" for someone to be present, recording observations for research and publication purposes (Thorne 1993: 20).

Participant observation: A qualitative method for gathering data that involves developing a sustained relationship with people while they go about their normal activities.

Complete observation: A role in participant observation in which the researcher does not participate in group activities and is publicly defined as a researcher.

Reactive effects: The changes in an individual or group behavior that result from being observed or otherwise studied.

Mixed Participation or Observation

Most field researchers adopt a role that involves some active participation in the setting. Usually they inform at least some group members of their research

interests, but then they participate in enough group activities to develop rapport with members and to gain a direct sense of what group members experience. This is not an easy balancing act. In his massive, 10-year study of gangs in urban America, Martin Sanchez Jankowski (1991: 13) participated in nearly all the things they did. “I ate where they ate, I slept where they slept, I stayed with their families, I traveled where they went, and . . . I fought with them. The only things that I did not participate in were those activities that were illegal . . . (including taking drugs).”

And Jankowski (1991) says that although, for instance, the fights he was in “often left bruises, I was never seriously hurt. Quite remarkably, in the more than 10 years during which I conducted this research, I was only seriously injured twice” (p. 12).

A strategy of mixed participation and observation has two clear ethical advantages. Because group members know the researcher’s real role in the group, they can choose to keep some information or attitudes hidden. By the same token, a researcher such as Jankowski can decline to participate in unethical or dangerous activities. Most field researchers get the feeling that, after they have become known and at least somewhat trusted figures in the group, their presence does not have any palpable effect on members’ actions.

Video Link

Watch a lecture on participant observation and ethnography.

One especially interesting example of a mixed strategy is Chambliss’s work on Olympic-level competitive swimmers. While working as a pure observer with a large number of world-class swimmers and teams, Chambliss himself coached, for 6 years, a small, local team in New York State. Here he tried to apply what he had learned through his years of research about what produces Olympic athletes. If his theories were correct, he reasoned, he should be able to make his own team much better. And, in fact, his swimmers improved dramatically, from being a rather poor local team to producing some state champions and even a few national-class athletes (Chambliss 1989). His written reports thus include a very unusual mix of observations, theorizing, and practical field experimentation to test his theory.

Complete Participation

Some field researchers adopt a **complete participation** role in which one operates as a fully functioning member of the setting. Most often, such research is also **covert**, or secret—other members don't know that the researcher is doing research. In one famous covert study, Laud Humphreys (1970) served as a "watch queen" so that he could learn about men engaging in homosexual acts in a public restroom. In another case, Randall Alfred (1976) joined a group of Satanists to investigate group members and their interaction. And Erving Goffman (1961) worked as a state mental hospital attendant while studying the treatment of psychiatric patients.

Covert participants don't disrupt their settings, but they do face other problems. They must write up notes from memory and must do so when it would be natural for them to be away from group members. Researchers often run to the bathroom to scribble their notes, jot reminders on napkins to expand on later, or whisper into hand-held recorders when they are out of the room. Researchers' spontaneous reactions to every event are unlikely to be consistent with those of the regular participants (Mitchell 1993), because they are not "really" interested in washroom sex, Satanists, or psychiatric ward attendants. When Diamond (1992) did covert research as an aide in a nursing home, his economic resources showed:

> "There's one thing I learned when I came to the States," [said a Haitian nursing assistant]. "Here you can't make it on just one job." She tilted her head, looked at me curiously, then asked, "You know, Tim, there's just one thing I don't understand about you. How do you make it on just one job?" (pp. 47–48)

Ethical issues have been at the forefront of the debate over the strategy of covert participation. Some covert observers may become so wrapped up in the role they are playing that they adopt not just the mannerisms but also the perspectives and goals of the regular participants—they "go native"—and so may end up "going along to get along" with group activities that are themselves unethical. Kai Erikson (1967) argues that covert participation is, therefore, by its very nature unethical and should not be allowed except in public settings. If others suspect the researcher's identity or if the researcher contributes to, or impedes, group

action, these consequences can be adverse. Covert researchers cannot anticipate the unintended consequences of their actions for research subjects or even for other researchers; covert research may, for instance, increase public distrust of all social scientists.

Complete (covert) participation: A role in field research in which the researcher does not reveal his or her identity as a researcher to those who are observed.

Entering the Field

Entering the field, the setting under investigation, is a critical stage in a participant observation project. Chambliss (1996) used a very "soft" technique for gaining access to hospitals. Rather than preparing a formal proposal to present to top administrators, he began quite informally:

> I use an informal series of contacts with lower level members of the organization. In the present study, I would try first to meet some staff nurses who worked at the target hospitals, see them socially—for instance, by inviting them to lunch—and tell them I was interested in learning about nursing, hospitals, and ethical problems therein. This gave me a chance, first, to learn a lot about nursing in a comfortable setting. More important, it gave the people I met a chance to see that I was easy to talk to, trustworthy, and a decent human being who was not out to do an exposé.
>
> Typically, such conversations ended with my new acquaintance suggesting that I talk with still another nurse or administrator and providing a phone number. I would immediately follow up on this suggestion. A series of such meetings and introductions typically concluded in my being invited by suitably authorized administrators to visit the hospital, observe various units, and talk with whomever I pleased. At that point, as needed, I would present a formal proposal for research, get necessary permission, and so on. Basically, my assumption is that once potential subjects get to know me, they won't be afraid of my doing research on them. (pp. 190–191)

When participant observing involves public figures who are used to reporters and researchers, a more direct approach may secure entry into the field. Richard Fenno (1978: 257) simply wrote a letter to most of the members of Congress whom he sought to study, asking for their permission to observe them at work. He received only two refusals and attributed this high rate of subject cooperation to such reasons as interest in a change in the daily routine, commitment to making themselves available, a desire for more publicity, the flattery of scholarly attention, and interest in helping to teach others about politics. Other groups have other motivations, but in every case, some consideration of these potential motives in advance should help smooth entry into the field.

In short, field researchers must be very sensitive to the impression they make and the ties they establish when entering the field. This stage lays the groundwork for collecting data from people who have different perspectives and for developing relationships that the researcher can use to surmount the problems in data collection that inevitably arise in the field. The researcher should be ready to explain to participants why he or she is involved in the field and how they might benefit from that involvement. Discussion about these issues with key participants, or gatekeepers, should be honest and should identify what the participants can expect from the research, without necessarily going into detail about the researcher's hypotheses or research questions (Rossman & Rallis 1998: 51–53, 105–108).

Gatekeeper: A person in a field setting who can grant researchers access to the setting.

Developing and Maintaining Relationships

Researchers must be careful to manage their relationships in the research setting so that they can continue to observe and interview diverse members of the setting throughout the long period typical of participant observation (Maxwell 1996: 66). Interaction early in the research process is particularly sensitive because participants don't know the researcher and the researcher doesn't know the group norms.

Journal Link

Read about how qualitative interviews allow the researcher to connect with participants.

In his classic *study Street Corner Society*, William F. Whyte (1955) used what in retrospect was a sophisticated two-part strategy to develop and maintain relationships with poor men whose informal relationships he studied in "Cornerville" (an Italian American slum neighborhood in Boston). The first part of Whyte's strategy was to maintain good relations with a group leader known as Doc and, through Doc, to stay on good terms with the others. Doc became a key informant in the research setting—a knowledgeable insider who knew the group's culture and was willing to share access and insights with the researcher (Gilchrist & Williams 1999). The less obvious part of Whyte's strategy was a consequence of his decision to move into Cornerville, a move he decided was necessary to understand and be accepted in the community fully. The room he rented in a local family's home became his base of operations. In some respects, this family became an important dimension of Whyte's immersion in the community: He tried to learn Italian by speaking with family members, and they conversed late at night as if Whyte were a real family member. But Whyte recognized that he needed a place to unwind after his days of constant alertness in the field, so he made a conscious decision not to include the family as an object of study. Living in this family's home became a means for Whyte to maintain standing as a community insider without becoming totally immersed in the demands of research (Whyte 1955: 294–297).

Experienced participant observers recommend developing a plausible (and

honest) explanation for yourself and your study and keeping the support of key individuals to maintain relationships in the field. They also suggest being somewhat laid-back, neither showing off your expertise nor being too aggressive in questioning others. Another good bit of advice is not faking social similarity with those you are observing and not offering monetary rewards for participation (Bogdewic 1999: 53–54; Rossman & Rallis 1998: 105–108; Whyte 1955: 300–306; Wolcott 1995: 91–95).

Key informant: An insider who is willing and able to provide a field researcher with superior access and information, including answers to questions that arise during the research.

Research That Matters

People can be very creative in trying to meet their basic needs after disaster strikes. Sociologist Yuki Kato at Tulane University and his collaborators Catarina Passidomo and Daina Harvey (2013) sought to understand how urban gardening, for instance, developed and became a political tool after Hurricane Katrina devastated New Orleans in 2005. Using participant observation, they conducted an ethnographic investigation of urban gardening projects that had the intentional, political goal of changing the allocation of resources in neighborhoods. They found that gardening projects ranged from the more political—"Our vision is to have the Lower Ninth Ward speak as one voice regarding what we want for food access in our neighbourhood"—to the less political—"Hollygrove Market and Farm exists to increase accessibility of fresh produce to Hollygrove"—but also found that priorities and politics shifted over time as the broader political climate changed.

Source: Adapted from Kato, Yuki, Catarina Passidomo, and Daina Harvey. 2014. Political gardening in post-disaster city: Lessons from New Orleans. *Urban Studies* 51:1833–1849.

Sampling People and Events

Qualitative researchers intensively study people, places, or other phenomena of interest, so they tend to limit their focus to just one or a few sites or programs. Still, the sample must be appropriate and adequate for the study, even if it is not representative. The qualitative researcher may select a *critical case* that is unusually rich in information pertaining to the research question; a *typical case*, precisely because it is judged to be typical; or a *deviant case*, which provides a useful contrast (Kuzel 1999). Within a research site, plans may be made to sample different settings, people, events, and artifacts (Exhibit 9.2).

Studying more than one case or setting almost always strengthens the causal conclusions and makes the findings more generalizable (King, Keohane, & Verba 1994). For example, Diamond (1992) worked in three different Chicago nursing homes "in widely different neighborhoods" and with different percentages of residents supported by Medicaid. He then "visited many homes across the United States to validate my observations" (p. 5).

Exhibit 9.2 Sampling Plan for Participant Observation in Schools

Information Source*	Type of Information to Be Obtained				
	Collegiality	Goals and Community	Action Expectations	Knowledge Orientation	Base
Settings					
Public places (halls, main offices)					
Teacher's lounge	X	X		X	X
Classrooms		X	X	X	X
Meeting rooms	X		X	X	
Gymnasium or locker room		X			
Events					
Faculty meetings	X		X		X
Lunch hour	X				X
Teaching		X	X	X	X
People					
Principal		X	X	X	X
Teachers	X	X	X	X	X
Students		X	X	X	
Artifacts					
Newspapers		X	X		X
Decorations		X			

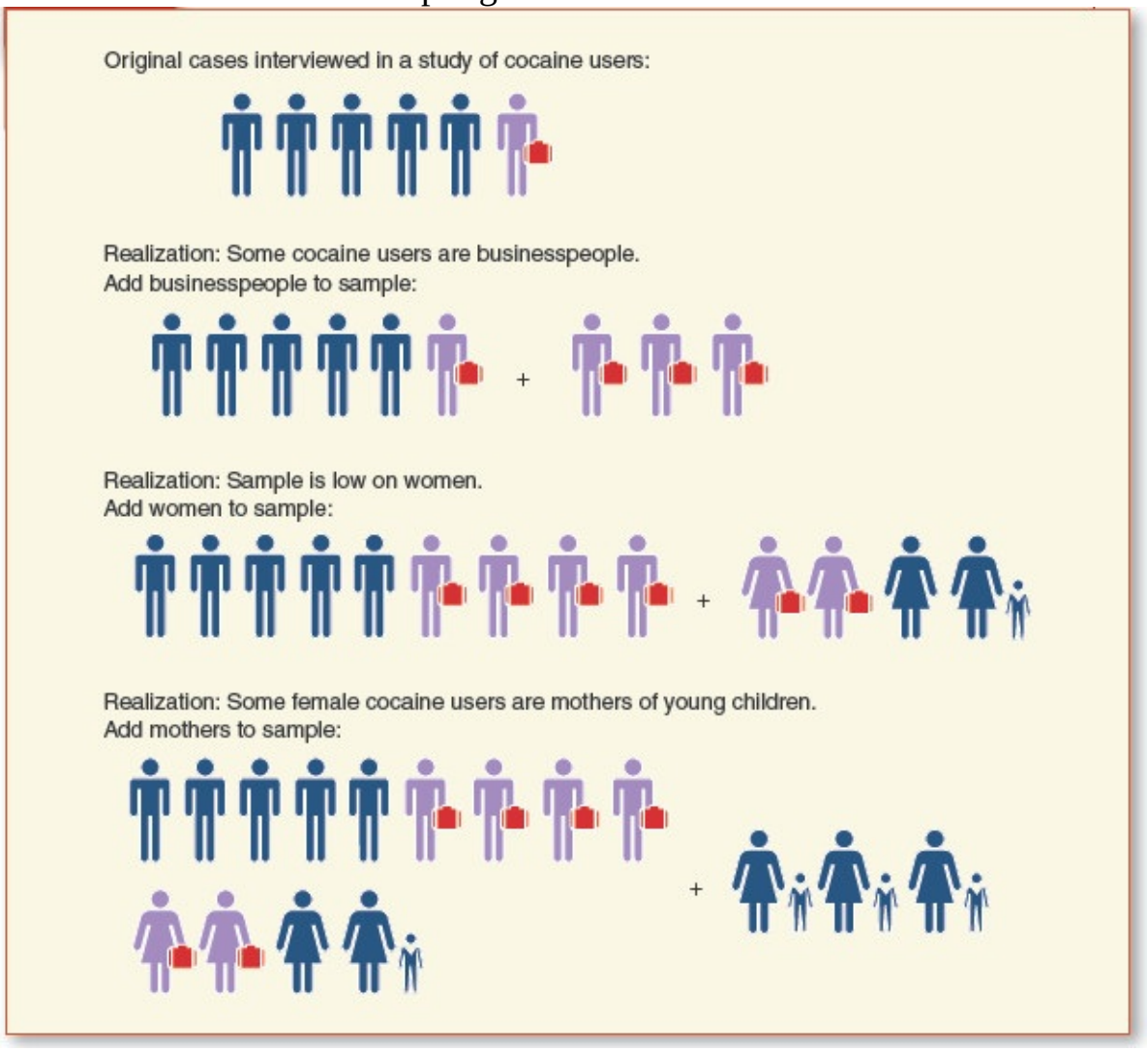
*Selected examples in each category.

Other approaches to sampling in field research are more systematic. Researchers use theoretical sampling when they focus their investigation on particular processes that seem to be important and select instances to allow comparisons or checks with which they can test these perceptions (Ragin 1994: 98–101) (Exhibit 9.3). Jankowski (1991), again, provides an impressive example of conscientious theoretical sampling in field research:

It was first essential to investigate *gangs in different cities* in order to control for the different socioeconomic and political environments that they operate in. Second, in order to determine if there were any differences associated with ethnicity, it was critical to *compare gangs composed of different ethnic groups.* Three metropolitan areas were therefore chosen for the study: the greater Los Angeles area, various boroughs of New York City, and the greater Boston area.

Two were eastern cities with certain weather patterns; the other was western with a completely different weather pattern. (Weather has often been thought to have an impact on gang activity, with colder weather restricting activity and warmer weather encouraging it.)

Exhibit 9.3 Theoretical Sampling


Original cases interviewed in a study of cocaine users:
Realization: Some cocaine users are businesspeople.
Add businesspeople to sample:
Realization: Sample is low on women.
Add women to sample:
Realization: Some female cocaine users are mothers of young children.
Add mothers to sample:

Of the thirty-seven gangs studied, thirteen were in the Los Angeles area, twenty were in the New York City area, and four were in the Boston area. Various ethnic groups are represented in the sample, which includes gangs composed of Irish, African-American, Puerto Rican, Chicano, Dominican, Jamaican, and Central American members. The sample also involves gangs of varying size. The smallest had thirty-four members; the largest had more than one thousand . . . Within this sample, stratified by ethnicity, I randomly selected ten in each city.

It was my intention to study African-American gangs, Latino gangs, Asian gangs, and white gangs, and so gangs representing each of these ethnic groups were chosen. Because I wanted to include gangs of varying membership sizes, I randomly selected gangs from my ethnically stratified list until I obtained a sample representing gangs of different sizes. Since my overall strategy was to study five gangs in Los Angeles and five in New York for two years, then add more, and finally add several Boston gangs, I selected five of the original ten chosen and began my effort to secure their participation. (Jankowski 1991: 6–7)

Theoretical sampling: A sampling method recommended for field researchers by Glaser and Strauss (1967). A theoretical sample is drawn in a sequential fashion, with settings or individuals selected for study as earlier observations or interviews indicate that these settings or individuals are influential.

Taking Notes

Notes are the primary means of recording participant observation data (Emerson, Fretz, & Shaw 1995). It is almost always a mistake to try to take comprehensive notes while engaged in the field—the process of writing extensively is just too disruptive. The usual procedure is to jot down brief notes about highlights of the observation period. These brief notes then serve as memory joggers when writing the actual **field notes** later. It also helps to maintain a daily log in which each day's activities are recorded (Bogdewic 1999: 58–67). With the aid of the **jottings** and some practice, researchers usually remember a great deal of what happened—as long as the comprehensive field notes are written immediately afterward or at least within the next 24 hours, and before they have been discussed with anyone else.

Usually writing up notes takes much longer—at least three times longer—than the observing did. Field notes must be as complete, detailed, and true to what was observed and heard as possible. Direct quotes should be distinguished clearly from paraphrased quotes, and both should be set off from the researcher's observations and reflections. The surrounding context should receive as much attention as possible, and a map of the setting should be included, with indications of where individuals were at different times.

Careful note taking yields a big payoff. On page after page, field notes will suggest new concepts, causal connections, and theoretical propositions. Notes also should include descriptions of the methodology and a record of the researcher's feelings and thoughts while observing. Exhibit 9.4 illustrates these techniques with notes from the Chambliss study.

Encyclopedia Link

Read about the importance of field notes in qualitative research.

Field notes: Notes that describe what has been observed, heard, or otherwise experienced in a participant observation study. These notes usually are written after the observational session.

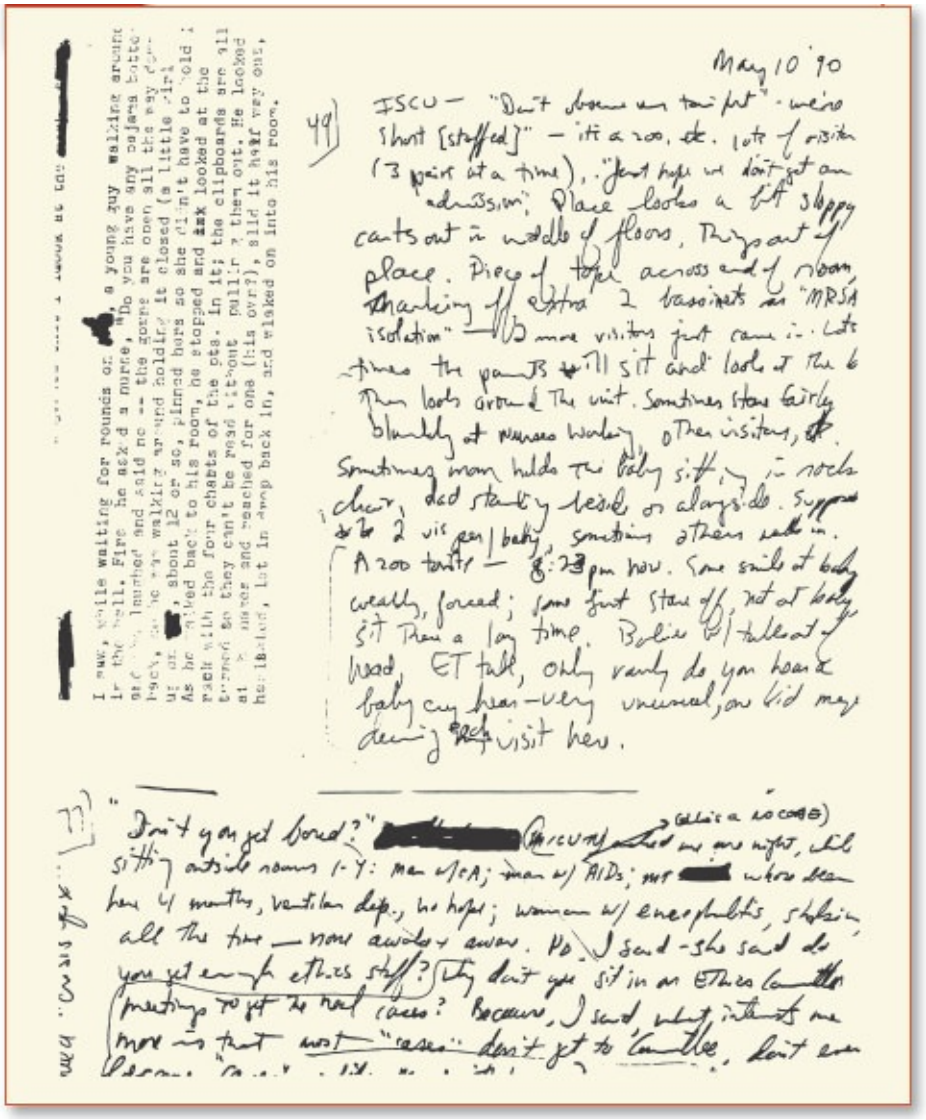
Jottings: Brief notes written in the field about highlights of an observation period.

Managing the Personal Dimensions

Field researchers cannot help but be affected on a personal, emotional level by social processes in the social situation they are studying. At the same time, those being studied react to researchers not just as researchers but as personal acquaintances—and often as friends. Managing and learning from this personal side of field research is an important part of any project.

The researcher, like his informants, is a social animal. He has a role to play, and he has his own personality needs that must be met in some degree if he is to function successfully. Where the researcher operates out of a university, just going into the field for a few hours at a time, he can keep his personal social life separate from field activity. His problem of role is not quite so complicated. If, on the other hand, the researcher is living for an extended period in the community she is studying, her personal life is inextricably mixed with her research (Whyte 1955: 279).

Exhibit 9.4 Sample Field Notes From the Chambliss Nursing Study



May 10 '90

49

ISCU — "Don't observe us tonight" - we're short [staffed]" — it a zoo, etc. lots of visits (3 pairs at a time), "Just hope we don't get an admission". Place looks a bit sloppy carts out in middle of floors. Thin out of place. Piece of tape across end of room, marking off extra 2 bassinets as "MRSA isolation" — 10 more visitors just came in. Lots times the parents will sit and look at the b then looks around the unit. Sometimes stare fairly blankly at nurses working. Other visitors, Sometimes mom holds the baby sitting in rocking chair, dad standing beside or alongside. Suppose to be 2 visitors/baby, sometimes others walk in. A zoo tonight — 8:23 pm now. Some smile at baby weakly, forced; some just stare off, not at baby sit there a long time. Babies bald at head, ET tube, only rarely do you hear baby cry here — very unusual, one kid may during visit here.

"Don't you get bored?" [redacted] (Marcus) (this is a no code) we are night, all sitting outside rooms 1-4: man w/ CA; man w/ AIDs; mom [redacted] whose been here 4 months, ventilator dep., no hope; woman w/ encephalitis, shaking all the time — more awake + aware. No. I said - she said do you get enough ethics stuff? Why don't you sit in on Ethics Committee meetings to get to real cases? Because, I said, what interests me more is that most "cases" don't get to Committee, don't even

Note: Original field notes, either written on site or typed later that day. Identifying information has been blacked out. “ISCU” stands for “Infant Special Care Unit,” where premature infants are cared for. The first sentence reads, “Don’t observe us tonight,” “we’re short [staffed],” a quotation from a nurse in the unit.

Barrie Thorne (1993), a sociologist known for her research on gender roles among children, wondered whether “my moments of remembering, the times when I felt like a ten-year-old girl, [were] a source of distortion or insight?” (p. 26). She concluded they were both: “Memory, like observing, is a way of knowing and can be a rich resource.” But “when my own responses . . . were

driven by emotions like envy or aversion, they clearly obscured my ability to grasp the full social situation" (p. 26).

There is no formula for successfully managing the personal dimension of field research. It is much more art than science and flows more from the researcher's own personality and natural approach to other people than from formal training. But novice field researchers often neglect to consider how they will manage personal relationships when they plan and carry out their projects. Attention to a few guidelines based on our personal experience with field research, provided in Exhibit 9.5, should maximize the likelihood of a project's success.

Systematic Observation

Observations can be made in a more systematic, quantitative design that allows systematic comparisons and more confident generalizations. A researcher using systematic observation develops a standard form on which to record variation within the observed setting for variables of interest. Such variables might include the frequency of some behavior(s), the particular people observed, the weather or other environmental conditions, and the number and state of repair of physical structures. In some systematic observation studies, records will be obtained from a random sample of places or times.

Robert Sampson and Stephen Raudenbush's (1999) study of disorder and crime in urban neighborhoods provides an excellent example of systematic observation methods. A systematic observational strategy increases reliability by using explicit rules that standardize coding practices across observers (Reiss 1971). It is a method particularly well suited to overcome one of the limitations of survey research on crime and disorder: Residents who are fearful of crime perceive more neighborhood disorder than do residents who are less fearful, even though both are observing the same neighborhood (Sampson & Raudenbush 1999: 606).

Exhibit 9.5 Nine Steps to Successful Field Research

1. ***Have a simple, one-sentence explanation of your project.*** "I want to learn about the problems nurses face in their work," or "I want to learn what makes a great swimming team." People will ask what you're doing, but no one cares to hear all your theories.
2. ***Be yourself.*** Don't lie about who you are. First, it's wrong. Second, you'll get caught and ruin the trust you're trying to build. (Yes, there are exceptions, but very few.)
3. ***Don't interfere.*** They got along just fine before you came along, and they can do it again. Don't be a pest.
4. ***Listen, actively.*** Be genuinely interested in what they say. Movie stars, politicians, and other celebrities are used to having other people listen to what they say, but that's not true for most people. If you really care to listen, they'll tell you everything.
5. ***Show up***, at every opportunity—3:00 in the morning, or if you have to walk 5 miles. Go to their parties and their funerals. Make a 5-hour trip for a 15-minute interview, and they'll notice—and give you everything you want.
6. ***Pay attention to everything***, especially when you're bored. That's when the important stuff is happening, the stuff ***no one else*** notices.
7. ***Protect your sources***, more than is necessary. When word gets around that you can be trusted, you won't believe what people will tell you.
8. ***Write everything down, that day.*** By tomorrow, you'll forget 90% of the best material, and then it's gone forever.
9. ***Always remember: It's not about you, it's about them.*** Don't try to be smart, or savvy, or hip; don't try to be the center of attention. Stop thinking about yourself all the time. Pay attention to other people.

Source: Raudenbush, Stephen W., and Robert J. Sampson. 1999. Econometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods. Sociological Methodology 29:1–41.

This ambitious multiple-methods investigation combined observational research, survey research, and archival research. The observational component involved a stratified probability (random) sample of 196 Chicago census tracts. A specially equipped sport-utility vehicle was driven down each street in these tracts at the rate of 5 miles per hour. Two video recorders taped the blocks on both sides of the street, while two observers peered out of the vehicle's windows and recorded their observations in the logs. The result was an observational record of 23,816 face blocks (the block on one side of the street is a face block). The observers recorded in their logs codes that indicated land use, traffic, physical conditions, and evidence of physical disorder (Exhibit 9.6). The videotapes were sampled and then coded for 126 variables, including housing characteristics, businesses, and social interactions. Physical disorder was measured by counting such features as cigarettes or cigars in the street, garbage, empty beer bottles, graffiti,

condoms, and syringes. Indicators of social disorder included adults loitering, drinking alcohol in public, fighting, and selling drugs. To check for reliability, a different set of coders recoded the videos for 10% of the blocks. The repeat codes achieved 98% agreement with the original codes.

Sampson and Raudenbush also measured crime levels with data from police records, census tract socioeconomic characteristics with census data, and resident attitudes and behavior with a survey. The combination of data from these sources allowed a test of the relative impact on the crime rate of residents' informal social control efforts and of the appearance of social and physical disorder.

Peter St. Jean (2007) extended the research of Sampson and Raudenbush with a mixed-methods study of high crime areas that used resident surveys, participant observation, in-depth interviews with residents and offenders, and systematic social observation. St. Jean recorded neighborhood physical and social appearances with video cameras mounted in a van that was driven along neighborhood streets. Pictures were then coded for the presence of neighborhood disorder (Exhibit 9.7).

This study illustrates both the value of multiple methods and the technique of recording observations in a form from which quantitative data can be obtained. The systematic observations give us much greater confidence in the measurement of relative neighborhood disorder than we would have from unstructured descriptive reports or from responses of residents to survey questions. Interviews with residents and participant observation helped to identify the reasons that offenders chose particular locations when deciding where to commit crimes.

How Do You Conduct Intensive Interviews?

Participant observation can provide a wonderfully rich view, then, of the social world. But it remains a *view*, seen by the observer. Often we wonder what individuals think or feel or how they see their world. For this purpose, one can use intensive interviews.

Unlike the more structured interviewing that may be used in survey research (discussed in Chapter 7), intensive, or depth, interviewing relies on open-ended questions to develop a comprehensive picture of the interviewee's background, attitudes, and actions—to "listen to people as they describe how they understand the worlds in which they live and work" (Rubin & Rubin 1995: 3).

For instance,

> We had two or three patients, and they were terminally ill with cancer. We would give the patients, every two or three hours around the clock toward the end, morphine sulfate intramuscular.
>
> I was really worried about giving them a morphine injection because the morphine depresses the respiration. I thought, well, is this injection going to do them in?
>
> If I don't give the injection, they will linger on longer, but they might also have more pain. If I do give the injection, the end result of death is going to occur faster. Am I playing God?" [Interview] (Chambliss 1996: 171)

Exhibit 9.6 Neighborhood Disorder Indicators Used in Systematic Observation Log

Variable	Category	Frequency
Physical Disorder		
Cigarettes, cigars on street or gutter	no	6,815
	yes	16,758
Garbage, litter on street or sidewalk	no	11,680
	yes	11,925
Empty beer bottles visible in street	no	17,653
	yes	5,870
Tagging graffiti	no	12,859
	yes	2,252
Graffiti painted over	no	13,390
	yes	1,721
Gang graffiti	no	14,138
	yes	973
Abandoned cars	no	22,782
	yes	806
Condoms on sidewalk	no	23,331
	yes	231
Needles/syringes on sidewalk	no	23,392
	yes	173
Political message graffiti	no	15,097
	yes	14
Social Disorder		
Adults loitering or congregating	no	14,250
	yes	861
People drinking alcohol	no	15,075
	yes	36
Peer group, gang indicators present	no	15,091
	yes	20
People intoxicated	no	15,093
	yes	18
Adults fighting or hostilely arguing	no	15,099
	yes	12
Prostitutes on street	no	15,100
	yes	11
People selling drugs	no	15,099
	yes	12

Source: St. Jean, Peter K. B. 2007. Pockets of crime: Broken windows, collective efficacy, and the criminal point of view. Chicago: University of Chicago Press.

Exhibit 9.7 One Building in St. Jean's (2007) Study

Source: St. Jean, Peter K. B. 2007. Pockets of crime: Broken windows, collective efficacy, and the criminal point of view. Chicago: University of Chicago Press. Reprinted with permission.

Audio Link

Listen to more information on conducting interviews.

The key to eliciting such a response is *active listening*—which is not the same as just being quiet. Instead, you must actively question, ask for explanations, and show a genuine deep curiosity about the subject's views and feelings. Your own opinions are not important here; you must suspend all judgment of what the respondent is saying, even if you regard the person's opinions as obnoxious or even immoral. Remember, the goal is to learn what the *respondent* thinks, not to

express what you think.

Therefore, depth interviews may be highly unstructured. Rather than asking standard questions in a fixed order, a researcher conducting intensive interviews may allow the specific content and order of questions to vary from one interviewee to another. Like participant observation studies, intensive interviewing engages researchers actively with subjects. The researchers must listen to lengthy explanations, ask follow-up questions tailored to the preceding answers, and seek to learn about interrelated belief systems or personal approaches to things rather than measure a limited set of variables. As a result, intensive interviews are often much longer than standardized interviews, sometimes as long as 15 hours, conducted in several different sessions.

The intensive interview can become more like a conversation between partners than between a researcher and a subject (Kaufman 1986: 22–23). Some call it "a conversation with a purpose" (Rossman & Rallis 1998: 126). Robert Bellah and his colleagues (1985) elaborate on this aspect of intensive interviewing in a methodological appendix to their national best seller about American individualism, *Habits of the Heart:*

> We did not, as in some scientific version of "Candid Camera," seek to capture their beliefs and actions without our subjects being aware of us. Rather, we sought to bring our preconceptions and questions into the conversation and to understand the answers we were receiving not only in terms of the language but also, so far as we could discover, in the lives of those we were talking with. Though we did not seek to impose our ideas on those with whom we talked . . . we did attempt to uncover assumptions, to make explicit what the person we were talking to might rather have left implicit. The interview as we employed it was active, Socratic. (p. 304)

Random selection is rarely used to select respondents for intensive interviews, but the selection method still must be considered carefully. Researchers should try to select interviewees who are knowledgeable about the subject of the interview, who are open to talking, and who represent a range of perspectives (Rubin & Rubin 1995: 65–92). Selection of new interviewees should continue, if possible, at least until the **saturation point** is reached, the point when new interviews seem to yield little additional information (Exhibit 9.8).

Establishing and Maintaining a Partnership

Because intensive interviewing does not engage researchers as participants in subjects' daily affairs, the problems of entering the field are much reduced. However, the logistics of arranging long periods for personal interviews can still be pretty complicated. It also is important to establish rapport with subjects by considering in advance how they will react to the interview arrangements and by developing an approach that does not violate their standards for social behavior. Interviewees should be treated with respect, as knowledgeable partners whose time is valued (in other words, don't be late for your appointments). A commitment to confidentiality should be stated and honored (Rubin & Rubin 1995).

Research|Social Impact Link

Read about how qualitative studies help create bodies of research and inform public policy.

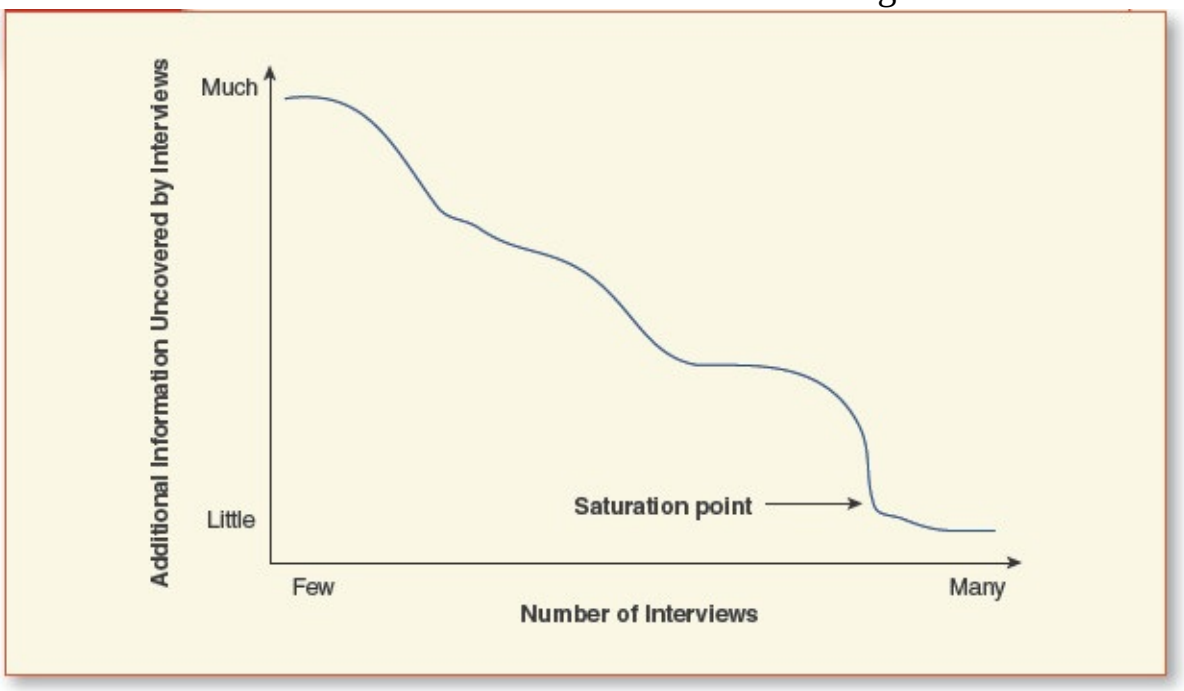
Intensive (depth) interviewing: A qualitative method that involves open-ended, relatively unstructured questioning in which the interviewer seeks in-depth information on the interviewee's feelings, experiences, and perceptions.

Saturation point: The point at which subject selection is ended in intensive interviewing because new interviews seem to yield little additional information.

Asking Questions and Recording Answers

Intensive interviewers must plan their main questions around an outline of the interview topic. The questions generally should be short and to the point. More details can then be elicited through nondirective probes (such as “Can you tell me more about that?” or “Uh-huh,” echoing the respondent’s comment, or just maintaining a moment of silence). Follow-up questions can then be tailored to answers to the main questions.

Exhibit 9.8 The Saturation Point in Intensive Interviewing



Interviewers should strategize throughout an interview about how best to achieve their objectives while considering interviewees’ answers. *Habits of the Heart* (Bellah et al. 1985) again provides a useful illustration:

[Coinvestigator Steven] Tipton, in interviewing Margaret Oldham [a pseudonym], tried to discover at what point she would take responsibility for another human being:

Q: So what are you responsible for?

A: I’m responsible for my acts and for what I do.

Q: Does that mean you're responsible for others, too?

A: No.

Q: Are you your sister's keeper?

A: No.

Q: Your brother's keeper?

A: No.

Q: Are you responsible for your husband?

A: I'm not. He makes his own decisions. He is his own person. He acts his own acts. I can agree with them, or I can disagree with them. If I ever find them nauseous enough, I have a responsibility to leave and not deal with it any more.

Q: What about children?

A: I . . . I would say I have a legal responsibility for them, but in a sense I think they in turn are responsible for their own acts. (p. 304)

Do you see how the interviewer actively encouraged the subject to *explain* what she meant by "responsibility"? This sort of active questioning undoubtedly did a better job of clarifying the interviewee's concept of responsibility than a fixed set of questions would have.

Audio recorders commonly are used for recording intensive interviews and focus group interviews. They do not inhibit most interviewees and are routinely ignored. Occasionally respondents are very concerned with their public image and may therefore speak "for the recorder," but such individuals are unlikely to speak frankly in any research interview. In any case, constant note taking during an interview prevents adequate displays of interest and is distracting.

Interviewing Online

Our social world now includes many connections initiated and maintained through e-mail and other forms of web-based communication, so it is only natural that qualitative interviewing has also moved online. Interviewing online can facilitate interviews with others who are separated by physical distance; it also is a means to conduct research with those who are only known through such online connections as a discussion group or an e-mail distribution list (James & Busher 2009: 14).

Online interviews can be either synchronous—in which the interviewer and interviewee exchange messages as in online chatting—or asynchronous—in which the interviewee can respond to the interviewer's questions whenever it is convenient, usually through e-mail. Both styles of online interviewing have advantages and disadvantages (James & Busher 2009: 13–16). Synchronous interviewing provides an experience more similar to an in-person interview, thus giving more of a sense of obtaining spontaneous reactions, but it requires careful attention to arrangements and is prone to interruptions. Asynchronous interviewing allows interviewees to provide more thoughtful and developed answers, but it may be difficult to maintain interest and engagement if the exchanges continue over many days. The online asynchronous interviewer should plan carefully how to build rapport as well as how to terminate the online relationship after the interview is concluded (King & Horrocks 2010: 86–93).

Whether a synchronous or asynchronous approach is used, online interviewing can facilitate the research process by creating a written record of the entire interaction without the need for typed transcripts. The relative anonymity of online communications can also encourage interviewees to be more open and honest about their feelings than they would be if interviewed in person (James & Busher 2009: 24–25). However, online interviewing lacks some of the most appealing elements of qualitative methods: The revealing subtleties of facial expression, intonation, and body language are lost, and the intimate rapport that a good intensive interviewer can develop in a face-to-face interview cannot be achieved. In addition, those who are being interviewed have much greater ability to present an identity that is completely removed from their in-person persona; for instance, basic characteristics such as age, gender, and physical location can be completely misrepresented.

How Do You Run Focus Groups?

Finally, for quick, emotionally resonant answers, **focus groups** can be the qualitative researcher's best friend. Long favored by advertisers, marketing researchers, and political consultants who want to see "what message pushes their buttons," focus groups are collections of unrelated individuals, convened by a researcher and then led in group discussion of a topic for 1 to 2 hours. The researcher asks specific questions and guides the discussion, but the resulting information is qualitative and relatively unstructured. Focus groups need not involve representative samples; instead, a few individuals are recruited for the group who have the time to participate, have some knowledge pertinent to the focus group topic, and share key characteristics with the target population. Throughout the Mellon Project on liberal arts education at Hamilton College, focus groups—of dean's list students, minority students, or study abroad participants, for instance—have been used to assess major problem areas in various programs rapidly and to develop areas for more systematic investigation.

Focus group research typically proceeds like this: The researcher convenes a series of groups, each including 7 to 10 people, for the discussions. Sometimes the groups are heterogeneous, with many dissimilar people (old and young, boss and employees, Democrats and Republicans); this can stimulate a broader array of opinions. But usually groups are, by design, homogeneous by categories one wants to compare. For instance, a business might run eight focus groups, four from the sales offices and four from service offices, to learn how these different functions see their customers. Or a college could run focus groups of freshmen and sophomores to learn about the different ways these groups approach course registration. It's generally best (though not always possible) to have group members be strangers so that personal relationships don't affect their answers, and it's crucial to avoid power differentials—no bosses with subordinates, teachers with students, or parents with their children. Such combinations will prevent open and honest opinion from emerging (Krueger & Casey 2000).

Once completed, focus group discussions are relatively easy to analyze: Just compare the responses, on each question, from one kind of group (say, salespeople) to responses for the same question by another kind of group (say, service representatives).

Richard Krueger (1988) provides a good example of a situation in which focus groups were used effectively:

> [A] University recently launched a $100 million fund drive. The key aspect of the drive was a film depicting science and research efforts. The film was shown in over two dozen focus groups of alumni, with surprising results to University officials. Alumni simply did not like the film and instead were more attracted to supporting undergraduate humanistic education. (pp. 33–37)

Audio Link

Listen to advice on how to conduct a focus group.

Focus group methods share with other field research techniques an emphasis on discovering unanticipated findings and exploring hidden meanings. Although weak in developing reliable, generalizable results (the strength of survey research), focus groups can be indispensable for developing hypotheses and survey questions, for investigating the meaning of survey results, and for quickly assessing the range of opinion about an issue. Exhibit 9.9 presents guidelines, derived from Krueger and Casey (2000) for running focus groups.

Focus groups: A qualitative method that involves unstructured group interviews in which the focus group leader actively encourages discussion among participants on the topics of interest.

Ethical Issues in Qualitative Research

Qualitative research can raise some complex ethical issues. No matter how hard the field researcher strives to study the social world naturally, leaving no traces, the very act of research itself imposes something "unnatural" on the situation. It is up to the researchers themselves to identify and take responsibility for the consequences of their involvement. Five main ethical issues arise:

1. *Voluntary participation*—Ensuring that subjects are participating in a study voluntarily is not often a problem with intensive interviewing and focus group research, but it is often a point of contention in participant observation studies. Few researchers or institutional review boards are willing to condone covert participation because it does not offer any way to ensure that participation by the subjects is voluntary. Even when the researcher's role is more open, interpreting the standard of voluntary participation still can be difficult. Should the requirement of voluntary participation apply equally to every member of an organization being observed? What if the manager consents, the workers are ambivalent, and the union says no?

2. *Subject well-being*—Before beginning a project, every field researcher should consider carefully how to avoid harm to subjects. It is not possible to avoid every theoretical possibility of harm or to be sure that a project will cause no adverse consequences whatsoever to any individual, but direct harm to the reputations or feelings of particular individuals should be avoided at all costs. The risk of such harm can be minimized by maintaining the confidentiality of research subjects and by not adversely affecting the course of events while engaged in a setting. Whyte (1955: 335–337) found himself regretting having recommended that a particular politician be allowed to speak to a social club he was observing because the speech led to serious dissension in the club and strains between Whyte and some club members.

3. *Identity disclosure*—Current ethical standards require informed consent of research subjects, and most would argue that this standard cannot be met in any meaningful way if researchers do not disclose fully their identity. But how much disclosure about the study is necessary, and how hard should researchers try to make sure that their research purposes are understood? In field research on Codependents Anonymous, Leslie Irvine (1998) found that the emphasis on

anonymity and the expectations for group discussion made it difficult for her to disclose her identity. Can a balance be struck between the disclosure of critical facts and a coherent research strategy?

4. *Confidentiality*—Field researchers normally use fictitious names for the characters in their reports, but doing so does not always guarantee confidentiality to their research subjects. In Chambliss's nursing book, reference to "the director of the medical center" might have identified that person, at least to other employees of the center who knew Chambliss did his research there. And anyone studying public figures or national leaders in a social movement must exercise special care because their own followers or enemies can privately recognize such people. Researchers should thus make every effort to expunge any possible identifying material from published information and to alter unimportant aspects of a description when necessary to prevent identity disclosure. In any case, no field research project should begin if some participants clearly will suffer serious harm by being identified in project publications.

5. *Online research*—The large number of discussion groups and bulletin boards on the Internet has stimulated much interest in conducting research such as that of Nick Fox and Chris Roberts (1999), who observed physicians' LISTSERVs in the United Kingdom. Such research can violate the principles of voluntary participation and identity disclosure when researchers participate in discussions and record and analyze text but do not identify themselves as researchers (Associated Press 2000).

Exhibit 9.9 Keys to Running Focus Groups

- A great moderator—Is neutral and genuinely respects the participants and is a great listener who can draw people out.
- Main questions—These ask what you really want to know, can be answered by participants, are clear and understandable to the participants, and provide useful answers.
- Participants—Are homogeneous by relevant category for comparisons, with no power differentials within the group.
- Sampling—Is purposeful, representing the entire range of responses, and is random within the pools meeting criteria. Ideally, participants in any group should be strangers to each other. Use reminders to attend with incentives.
- Recording—Audio recording, with an assistant taking notes, is best.
- Analysis—Compare answers of different groups to different questions (groups on differently colored paper, sorted by question, etc.).
- Reporting—You are speaking for the participants. Lead with the big insights and answer the questions that were asked of the study. Interesting quotations get attention!
- When in doubt—Ask the potential participants about food, setting, issues, moderator, etc.

Basically, good focus groups get honest answers, on important topics, from people who know.

Source: Adapted from Krueger, Richard A., and Mary Anne Casey. 2000. Focus groups: A practical guide for applied research, 3rd ed. Thousand Oaks, CA: Sage. Copyright SAGE Publications. Used with permission.

These ethical issues cannot be evaluated independently. The final decision to proceed must be made after weighing the relative benefits and risks to participants. Few qualitative research projects will be barred by consideration of these ethical issues, however, except for those involving covert participation. The more important concern for researchers is to identify the ethically troublesome aspects of their proposed research and resolve them before the project begins, as well as to act on new ethical issues as they come up during the project.

Interactive Exercises

Qualitative Research

Conclusion

Qualitative research has both immediate and lasting attractions. Many of the classic works of social science, from Sigmund Freud's *Interpretation of Dreams* (1900/1999) and Margaret Mead's *Coming of Age in Samoa* (1928/2001) to Erving Goffman's *Presentation of Self in Everyday Life* (1959) and Kristin Luker's *Abortion and the Politics of Motherhood* (1985), rest on qualitative forms of social research. Telling true stories of real people, laying out their feelings and emotions, is qualitative research—interviews, fieldwork, and focus groups cut through the dry numbers and correlations, the abstract variables, and the hypotheses of contemporary quantitative social science. Qualitative research aims to go, as we said at the beginning of this chapter, where real people live. It thereby can become, at its best, a form of literature, beautifully teaching its readers the deeper truths of the human condition. More modestly, many students simply find reading reports of qualitative research to be far more interesting than the statistics used in survey analysis.

But "interesting" is not always the same as accurate, correct, or even representative. The juiciest stories that Chambliss heard from his nurses were not, as it happens, what typically happened in their lives. Researchers love a good quote, but it may not represent the truth of a setting; fieldworkers love finding a key informant whose views may not be those of the average subject. Like journalists, even the best qualitative researchers may be drawn to the odd, the unusual, or the available—and all of those may be poor substitutes for representative sampling, standardized questions, and other more sober approaches to learning about social life. The statistics of survey analysis and the control groups of experiments force us to face reality with self-discipline; they make it harder to fool ourselves about what we see.

In the end, qualitative methods are one—and only one—excellent set of tools, complementary in purpose to the tools of surveys, experiments, and other methods. Each has its strengths and its weaknesses. When surveys find that college students complain about "social life" but also rejoice that they "made my best friends ever here," interviews can explain the (apparent) contradiction. When police statistics and crime surveys can't fathom the logic of gang life, Martin Sanchez Jankowski steps in and tells us the story in all its richness. And remember: No experiment, however carefully designed with an eye to protecting

internal validity, could ever have uncovered what Sigmund Freud found by just sitting quietly next to a patient on a couch—and listening.

Key Terms

Complete observation 206
Complete (covert) participation 207
Ethnography 202
Ethnomethodology 205
Field notes 212
Field research 200
Focus groups 220
Gatekeeper 208
Intensive (depth) interviewing 215
Jottings 212
Key informant 209
Netnography (cyberethnography and virtual enthnography) 204
Participant observation 205
Qualitative methods 200
Reactive effects 206
Saturation point 218
Theoretical sampling 211

Highlights

- Qualitative methods are most useful in exploring new issues, in investigating hard-to-study groups, and in determining the meaning people give to their lives and actions. In addition, most social research projects can be improved in some respects by taking advantage of qualitative techniques.
- Ethnography involves immersion in a group or social setting to understand its culture, whereas netnography uses this process in research on online groups or social networks. Ethnomethodology studies the way that participants construct the social world in which they live.
- Qualitative researchers tend to develop ideas inductively; they try to understand the social context and sequential nature of attitudes and actions and explore the subjective meanings that participants attach to events. They rely primarily on participant observation, intensive interviewing, and, in recent years, focus groups.
- Participant observers may adopt one of several roles for a particular research project. Each role represents a different balance between observing and participating. Many field researchers prefer a moderate role, participating as well as observing in a group but acknowledging publicly the researcher role. Such a role avoids the ethical issues posed by covert participation while still allowing the insights into the social world derived from participating directly in it. The role that the participant observer chooses should be based on an evaluation of the problems likely to arise from reactive effects and the ethical dilemmas of covert participation.
- Field researchers must develop strategies for entering the field, developing and maintaining relations in the field, sampling, and recording and analyzing data. Selection of sites or other units to study may reflect an emphasis on typical cases, deviant cases, or critical cases that can provide more information than others. Sampling techniques commonly used within sites or in selecting interviewees in field research include theoretical sampling.
- Recording and analyzing notes is a crucial step in field research. Jottings are used as brief reminders about events in the field, whereas daily logs are useful to chronicle the researcher’s activities. Detailed field notes should be recorded daily. Periodic analysis of the notes can guide refinement of methods used in the field and of the concepts, indicators, and models

developed to explain what has been observed.

- Intensive interviews involve open-ended questions and follow-up probes, with the specific question content and order varying from one interview to another.
- Focus groups combine elements of participant observation and intensive interviewing. They can increase the validity of attitude measurement by revealing what people say when presenting their opinions in a group context instead of the artificial one-on-one interview setting.
- Computer software is used increasingly for the analysis of qualitative, textual, and pictorial data. Users can record their notes, categorize observations, specify links between categories, and count occurrences.
- The four main ethical issues in field research concern voluntary participation, subject well-being, identity disclosure, and confidentiality.

Student Study Site

Related Titles in Research Methods and Statistics From SAGE

Investigating the Social World: The Process and Practice of Research, Eighth Edition, by Russell K. Schutt

The Practice of Research in Social Work, Third Edition, by Rafael J. Engel and Russell K. Schutt

The Practice of Research in Criminology and Criminal Justice, Fifth Edition, by Ronet K. Bachman and Russell K. Schutt

Fundamentals of Research in Criminology and Criminal Justice, Third Edition, by Ronet K. Bachman and Russell K. Schutt

Designing Surveys: A Guide to Decisions and Procedures, Second Edition, by Ronald F. Czaja and Johnny Blair

A Guide to Qualitative Field Research, Second Edition, by Carol A. Bailey

Adventures in Social Research, Seventh Edition, by Earl Babbie, Fred Halley, William E. Wagner, III, and Jeanne Zaino

Adventures in Criminal Justice Research, Fourth Edition, by Kim Logio, George Dowdall, Earl Babbie, and Fred Halley

Social Statistics for a Diverse Society, Sixth Edition, by Chava Frankfort-Nachmias and Anna Leon-Guerrero

Essentials of Social Statistics for a Diverse Society by Anna Leon-Guerrero and Chava Frankfort-Nachmias

Research Methods in Psychology, Second Edition, by Paul G. Nestor and Russell K. Schutt

Research Methods in Education by Joseph Check and Russell K. Schutt

Fundamentals of Social Work Research, Second Edition, by Rafael J. Engel and Russell K. Schutt

Constructing Social Research by Charles C. Ragin and Lisa M. Amoroso

Using IBM® SPSS® Statistics for Social Statistics and Research Methods, Third Edition, by William E. Wagner, III

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore | Boston

The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Maurice Punch (1994) once opined that “the crux of the matter is that some deception, passive or active, enables you to get at data not obtainable by other means” (p. 91). What aspects of the social world would be difficult for participant observers to study without being covert? Might any situations require the use of covert observation to gain access? What might you do as a participant observer to lessen access problems while still acknowledging your role as a researcher?
2. Review the experiments and surveys described in previous chapters. Pick one and propose a field research design that would focus on the same research question but use participant observation techniques in a local setting. Propose the role that you would play in the setting, along the participant observation continuum, and explain why you would favor this role. Describe the stages of your field research study, including your plans for entering the field, developing and maintaining relationships, sampling, and recording and analyzing data. Then discuss what you would expect your study to add to the findings resulting from the study described in the book.
3. Intensive interviews are the core of many qualitative research designs. How do they differ from the structured survey procedures that you studied in the last chapter? What are their advantages and disadvantages over standardized interviewing? How does intensive interviewing differ from the qualitative method of participant observation? What are the advantages and disadvantages of these two methods?

Finding Research

1. Go to the *Annual Review of Sociology*'s website (http://annualreviews.org). Search for articles that use qualitative methods as the primary method of gathering data on any one of the following subjects: child development/socialization, gender/sex roles, or aging/gerontology. Enter "Qualitative AND Methods" in the subject field to begin this search. Review at least five articles, and report on the specific method of field research used in each.
2. Go to Intute's database of social sciences Internet resources (www.intute.ac.uk/socialsciences/). Choose "Research Tools and Methods" and then "Qualitative Methods." Now choose three or four interesting sites to find out more about field research—either professional organizations of field researchers or journals that publish their work. Explore the sites to find out what information they provide regarding field research, what kinds of projects are being done that involve field research, and the purposes for which specific field research methods are being used.
3. You have been asked to do field research on the World Wide Web's impact on the socialization of children in today's world. The first part of the project involves your writing a compare-and-contrast report on the differences between how you and your generation were socialized as children and the way children today are being socialized. Collect your data by surfing the web "as if you were a kid." The web is your field, and you are the field researcher.
4. Using any of the major search engines, explore the web within the "Kids" or "Children" subject heading, keeping field notes on what you observe.
5. Write a brief report based on the data you have collected. How has the web affected child socialization compared with when you were a child?

Critiquing Research

1. Read and summarize one of the qualitative studies discussed in this chapter or another classic study recommended by your instructor. Review and critique the study using the article review questions presented in Exhibit 12.2 on page 261. What questions are answered by the study? What questions are raised for further investigation?
2. 2. Write a short critique of the ethics of Carolyn Ellis's (1986) study (discussed in Chapter 2). Read the book ahead of time to clarify the details, and then focus on each of the ethical guidelines presented in this chapter: voluntary participation, subject well-being, identity disclosure, and confidentiality. Conclude with a statement about the extent to which field researchers should be required to disclose their identities and the circumstances in which they should not be permitted to participate actively in the social life they study.

Doing Research

1. Conduct a brief observational study in a public location on campus where students congregate. A cafeteria, a building lobby, or a lounge would be ideal. You can sit and observe, taking occasional notes unobtrusively and without violating any expectations of privacy. Observe for 30 minutes. Write up field notes, being sure to include a description of the setting and a commentary on your own behavior and your reactions to what you observed.
2. 2. Review the experiments and surveys described in previous chapters. Pick one and propose a field research design that would focus on the same research question but with participant observation techniques in a local setting. Propose the role along the participant observation continuum that you would play in the setting, and explain why you would favor this role. Describe the stages of your field research study, including your plans for entering the field, developing and maintaining relationships, sampling, and recording and analyzing data. Then discuss what you would expect your study to add to the findings resulting from the study described in the book.
3. Develop an interview guide that focuses on a research question addressed in one of the studies in this book. Using this guide, conduct an intensive interview with one person who is involved with the topic in some way. Take only brief notes during the interview; then write up as complete a record of the interview as you can immediately afterward. Turn in an evaluation of your performance as an interviewer and note taker together with your notes.

Ethics Questions

1. Should covert observation ever be allowed in social science research? Do you believe that social scientists should simply avoid conducting research on groups or individuals who refuse to admit researchers into their lives? Some have argued that members of privileged groups do not need to be protected from covert research by social scientists—that this restriction should only apply to disadvantaged groups and individuals. Do you agree? Why or why not?
2. Should any requirements be imposed on researchers who seek to study other cultures to ensure that procedures are appropriate and interpretations are culturally sensitive? What practices would you suggest for cross-cultural researchers to ensure that ethical guidelines are followed? (Consider the wording of consent forms and the procedures for gaining voluntary cooperation.)

Video Interview Questions

Listen to the researcher interviews for Chapter 9 at edge.sagepub.com/chamblissmssw5e.

1. What type of research design did Andrea Leverentz use in her study? What were some of the advantages and disadvantages of this type of design that were mentioned in the interview?
2. What new questions and issues came up during Leverentz's research, and how did these differ from the original research question or focus? What does this say about the inductive approach and the importance of, as Leverentz says, letting "the data speak to you"?
3. According to Lakshmi Srinivas, what are the benefits to ethnographic research?
4. What challenges of ethnographic research does Srinivas highlight?

Chapter 10 Qualitative Data Analysis

©iStockphoto.com/adempercem

Learning Objectives

1. Explain the meaning of an emic focus and of an etic focus in research and their relevance in qualitative data analysis.
2. Compare and contrast the use of narrative analysis and conversation analysis.
3. Describe the grounded theory approach and its role in data collection.
4. Identify changes in the social world that have led to the growth of visual sociology.
5. Give an example of the value of using more than one method of analyzing qualitative data in a project.
6. Discuss the ways in which computer-aided qualitative data analysis can facilitate research.
7. List three ethical issues that should be given special attention in qualitative data analysis.

> I was at lunch standing in line and he [another male student] came up to my face and started saying stuff and then he pushed me. I said . . . I'm cool with you, I'm your friend and then he push me again and calling me names. I told *him* to stop pushing me and then he push me hard and said something about my mom. And then he hit me, and I hit him back. After he fell I started kicking him.
>
> —Calvin Morrill et al. (2000: 521)

A real student writing an in-class essay about conflicts in which he had participated made this statement. It was written for a team of social scientists who were studying conflicts in high schools to better understand their origins and to inform prevention policies.

In qualitative data analysis, the raw data to be analyzed are text—words—rather than numbers. In the high school conflict study by Calvin Morrill and his colleagues (2000), there were initially no variables or hypotheses. The use of text, not numbers, and the (initial) absence of variables are just two of the ways in which qualitative analysis differs from quantitative.

In this chapter, we present and illustrate the features that most qualitative analyses share. There is no one correct way to analyze textual data. To quote

Michael Quinn Patton (2002), “Qualitative analysis transforms data into findings. No formula exists for that transformation. Guidance, yes. But no recipe. Direction can and will be offered, but the final destination remains unique for each inquirer, known only when—and if—arrived at” (p. 432).

We first discuss different types of qualitative analyses and then describe computer programs for qualitative data analysis. You will see that these increasingly popular programs are blurring the distinctions between quantitative and qualitative approaches to textual analysis.

What Is Distinctive About Qualitative Data Analysis?

The focus on text—on qualitative data rather than on numbers—is the most important feature of **qualitative data analysis**. The "text" that qualitative researchers analyze is most often transcripts of interviews or notes from participant observation sessions, but the term can also refer to pictures or other images that the researcher examines.

What can one learn from a text? There are two kinds of answers to this question. Some researchers view textual analysis as a way to understand what participants "really" thought or felt or did in some situation or at some point in time. The text becomes a way to get "behind the numbers" that are recorded in a quantitative analysis to see the richness of real social experience. In this approach, interviews or field studies can, for instance, illuminate what survey respondents really meant by their answers.

Other qualitative researchers, however, have adopted a *hermeneutic* perspective on texts, viewing interpretations as never totally true or false. The text has many possible interpretations (Patton 2002: 114). The meaning of a text, then, is negotiated among a community of interpreters, and to the extent that some agreement is reached about meaning at a particular time and place, that meaning can only be based on consensual community validation. From the hermeneutic perspective, a researcher constructs a "reality" with his interpretations of a text provided by the subjects of research; other researchers with different backgrounds could come to markedly different conclusions.

Qualitative and quantitative data analyses, then, differ in the priority given to the views of the subjects of the research versus those of the researcher. Qualitative data analysts seek to capture the setting or people who produced this text on their own terms rather than in terms of predefined (by researchers) measures and hypotheses. So, qualitative data analysis tends typically to be inductive—the analyst identifies important categories in the data, as well as patterns and relationships, through a process of discovery. There are often no predefined measures or hypotheses. Anthropologists term this an **emic focus**, which means representing the setting in terms of the participants, rather than an **etic focus**, in which the setting and its participants are represented in terms that the researcher brings to the study.

Good qualitative data analyses focus on the interrelated aspects of the setting or group, or person, under investigation—the case—rather than breaking the whole up into separate parts. The whole is always understood to be greater than the sum of its parts, so the social context of events, thoughts, and actions becomes essential for interpretation. Within this framework, it doesn't really make sense to focus on two variables out of an interacting set of influences and test the relationship between just those two.

Qualitative data analysis is an iterative and reflexive process that begins as data are being collected rather than after data collection has ceased (Stake 1995). Next to her field notes or interview transcripts, the qualitative analyst jots down ideas about the meaning of the text and how it might relate to other issues. This process of reading through the data and interpreting it continues throughout the project. When it appears that additional concepts need to be investigated or new relationships explored, the analyst adjusts the data collection. This process is termed **progressive focusing** (Parlett & Hamilton 1976).

> We emphasize placing an interpreter in the field to observe the workings of the case, one who records objectively what is happening but simultaneously examines its meaning and redirects observation to refine or substantiate those meanings. Initial research questions may be modified or even replaced in mid-study by the case researcher. The aim is to thoroughly understand [the case]. If early questions are not working, if new issues become apparent, the design is changed. (Stake 1995: 9)

Elijah Anderson (2003) describes the progressive focusing process in his memoir about his study of Jelly's Bar:

> I also wrote conceptual memos to myself to help me sort out my findings. Usually not more than a page long, they represented theoretical insights that emerged from my engagement with the data in my field notes. As I gained tenable hypotheses and propositions, I began to listen and observe selectively, focusing in on those events that I thought might bring me alive to my research interests and concerns. This method of dealing with the information I was receiving amounted to a kind of dialogue with the data, sifting out ideas, weighing new notions against the reality with which I [was] faced there on the streets and back at my desk. (pp. 235–236)

Following a few guidelines will help when a researcher starts analyzing qualitative data (Miller & Crabtree 1999):

- Know yourself—your biases and preconceptions.
- Know your question.
- Seek creative abundance. Consult others and keep looking for alternative interpretations.
- Be flexible.
- Exhaust the data. Try to account for all the data in the texts, then publicly acknowledge the unexplained and remember the next principle.
- Celebrate anomalies. They are the windows to insight.
- Get critical feedback. The solo analyst is a great danger to self and others.
- Be explicit. Share the details with yourself, your team members, and your audiences. (pp. 142–143)

Qualitative Data Analysis as an Art

If you miss the certainty of predefined measures and deductively derived hypotheses, you are beginning to understand the difference between quantitative and qualitative data analyses. Qualitative data analysis is even described by some as involving as much “art” as science—as a “dance.” In the words of William Miller and Benjamin Crabtree (1999),

> Interpretation is a complex and dynamic craft, with as much creative artistry as technical exactitude, and it requires an abundance of patient plodding, fortitude, and discipline. There are many changing rhythms; multiple steps; moments of jubilation, revelation, and exasperation. . . . The dance of interpretation is a dance for two, but those two are often multiple and frequently changing, and there is always an audience, even if it is not always visible. Two dancers are the interpreters and the texts. (pp. 138–139)

The “dance” of qualitative data analysis captures the alternation between immersion in the text to identify meanings and editing the text to create categories and codes. The process involves three steps in reading the text:

1. When the researcher reads the text literally, he or she is focused on its literal content and form; the text “leads” the dance.
2. Then the researcher reads the text reflexively, focusing on how his or her own orientation shapes interpretations and focus. Now, the researcher leads the dance.
3. Finally, the researcher reads the text interpretively; the researcher tries to construct his or her own interpretation of what the text means. (Miller & Crabtree 1999: 138–139)

Video Link

Watch and compare qualitative and quantitative data.

In this artful way, analyzing text involves both inductive and deductive

processes: The researcher generates concepts and linkages between them based on reading the text and checks the text to see whether his concepts and interpretations are reflected in it.

Qualitative data analysis: Techniques used to search and code textual, aural, and pictorial data and to explore relationships among the resulting categories.

Emic focus: Representing a setting with the participants' terms.

Etic focus: Representing a setting with the researcher's terms.

Progressive focusing: The process by which a qualitative analyst interacts with the data and gradually refines his or her focus.

Qualitative Compared With Quantitative Data Analysis

With these points in mind, let's review the differences of the logic behind qualitative versus quantitative analysis. Qualitative data analysis has the following characteristics (Denzin & Lincoln 2000: 8–10; Patton 2002: 13–14):

- A focus on meanings rather than on quantifiable phenomena
- Collection of much data on a few cases rather than little data on many cases
- Study in depth and detail, without predetermined categories or directions, rather than emphasis on analyses and categories determined in advance
- Conception of the researcher as an "instrument" rather than as the designer of objective instruments to measure particular variables
- Sensitivity to context, rather than seeking universal generalizations
- Attention to the impact of the researcher's and others' values on the course of the analysis, rather than presuming the possibility of value-free inquiry
- A goal of rich descriptions of the world rather than measurement of specific variables

Of course, even the most qualitative textual data can also be transposed to quantitative data through a process of categorization and counting. Some qualitative analysts also share with quantitative researchers a positivist goal of describing the world as it "really" is, but others have adopted a postmodern hermeneutic goal of trying to understand how different people see and make sense of the world, without believing that there is one uniquely correct description.

What Techniques Do Qualitative Data Analysts Use?

Most approaches to qualitative data analysis take five steps:

1. Documentation of the data and data collection
2. Conceptualization and coding
3. Examining relationships to show how one concept may influence another
4. Authenticating conclusions by evaluating alternative explanations, disconfirming evidence, and searching for negative cases
5. Reflexivity

The analysis of qualitative research notes begins in the field at the time of observation or interviewing, as the researcher identifies problems and concepts that appear likely to help in understanding the situation. Simply reading the notes or transcripts is an important step in the analytic process. Researchers should make frequent notes in the margins to identify important statements and to propose ways of coding the data: “husband/wife conflict,” perhaps, or “tension reduction strategy.”

An interim stage may consist of listing the concepts developed in the notes and perhaps diagramming the relationships among concepts (Maxwell 1996: 78–81). In large projects, regular team meetings are an important part of this process. In her study of neighborhood police officers, Susan Miller’s (1999) research team met to go over their field notes and to resolve points of confusion, as well as to talk with other skilled researchers who helped identify emerging concepts:

> The fieldwork team met weekly to talk about situations that were unclear and to troubleshoot any problems. We also made use of peer-debriefing techniques. Here, multiple colleagues, who were familiar with qualitative data analysis but not involved in our research, participated in preliminary analysis of our findings. (p. 233)

The back-and-forth of refining concepts usually continues throughout the entire qualitative research project.

Let’s examine each of the steps of qualitative analysis in more detail.

Documentation

The data for a qualitative study most often are notes jotted down in the field or during an interview or text transcribed from audiotapes. “The basic data are these observations and conversations, the actual words of people reproduced to the best of my ability from the field notes” (Diamond 1992: 7). What to do with all this material? As mentioned in Chapter 9, many novice researchers have become overwhelmed by the quantity of information, and their research projects have ground to a halt as a result.

Analysis is less daunting, however, if the researcher maintains a disciplined transcription schedule:

> Usually, I wrote these notes immediately after spending time in the setting or the next day. Through the exercise of writing up my field notes, with attention to “who” the speakers and actors were, I became aware of the nature of certain social relationships and their positional arrangements within the peer group. (Anderson 2003: 235)
>
> You can see Anderson’s analysis already emerging from the simple process of taking notes.

The first formal analytical step is documentation. The various contacts, interviews, written documents, and notes all need to be saved and catalogued in some fashion. Documentation is critical to qualitative research for several reasons: It is essential for keeping track of what will be a rapidly growing volume of notes, tapes, and documents; it provides a way of developing an outline for the analytic process; and it encourages ongoing conceptualizing and strategizing about the text.

Matthew Miles and A. Michael Huberman (1994: 53) provide a good example of a contact summary form that was used to keep track of observational sessions in a qualitative study of a new school curriculum (Exhibit 10.1).

Conceptualization, Coding, and Categorizing

Identifying and refining important concepts is a key part of the iterative process of qualitative research. Sometimes conceptualization begins with a simple observation that is interpreted directly, "pulled apart," and then put back together more meaningfully. Robert Stake provides an example (1995):

Exhibit 10.1 Example of a Contact Summary Form

Contact type:

Visit X _____

Phone _____

(with whom)

Site: Tindale

Contact date: 11/28-29/79

Today's date: 12/28/79

Written by: BLT

1. What were the main issues or themes that struck you in this contact?

   Interplay between highly prescriptive, "teacher-proof" curriculum that is top-down imposed and the actual writing of the curriculum by the teachers themselves.

   Split between the "watchdogs" (administrators) and the "house masters" (dept. chairs & teachers) vis-à-vis job foci.

   District curric, coord'r as decision maker re school's acceptance of research relationship.

2. Summarize the information you got (or failed to get) on each of the target questions you had for this contact.

Question	Information
History of dev. of innov'n	Conceptualized by Curric., Coord'r, English Chairman & Assoc. Chairman; written by teachers in summer; revised by teachers following summer with field testing data
School's org'l structure	Principal & admin'rs responsible for discipline; dept chairs are educ'l leaders
Demographics	Racial conflicts in late 60s; 60% black stud. pop.; heavy emphasis on discipline & on keeping out non-district students slipping in from Chicago
Teachers' response to innov'n	Rigid, structured, etc. at first; now, they say they like it/ NEEDS EXPLORATION
Research access	Very good; only restriction: teachers not required to cooperate

3. Anything else that struck you as salient, interesting, illuminating or important in this contact?

   Thoroughness of the innov'n's development and training.

   Its embeddedness in the district's curriculum, as planned and executed by the district curriculum coordinator.

   The initial resistance to its high prescriptiveness (as reported by users) as contrasted with their current acceptance and approval of it (again, as reported by users).

4. What new (or remaining) target questions do you have in considering the next contact with this site?

   How do users really perceive the innov'n? If they do indeed embrace it, what accounts for the change from early resistance?

   Nature and amount of networking among users of innov'n.

   Information on "stubborn" math teachers whose ideas weren't heard initially—who are they? Situation particulars? Resolution?

   Follow-up on English teacher Reilly's "fall from the chairmanship."

   Follow a team through a day of rotation, planning, etc.

   CONCERN: The consequences of eating school cafeteria food two days per week for the next four or five months . . .

Stop

Source: Miles, Matthew B., and A. Michael Huberman. 1994. Qualitative data analysis, 2nd ed. Thousand Oaks, CA: Sage. Used with permission.

When Adam ran a pushbroom into the feet of the children nearby, I jumped to conclusions about his interactions with other children: aggressive, teasing, arresting. Of course, just a few minutes earlier I had seen him block the children climbing the steps in a similar moment of smiling bombast. So I was aggregating, and testing my unrealized hypotheses about what kind of kid he was, not postponing my interpreting. . . . My disposition was to keep my eyes on him. (p. 74)

Encyclopedia Link

Read about the processes and strategies of qualitative data coding.

The focus in this conceptualization “on the fly” is to provide a detailed description of what was observed and a sense of why it was important.

More often, analytic insights are tested against new observations; the initial statement of problems and concepts is refined; and the researcher then collects more data, interacts with it again, and the process continues. Anderson (2003) recounts how his conceptualization of social stratification at Jelly’s Bar developed over a long period:

I could see the social pyramid, how certain guys would group themselves and say in effect, “I’m here and you’re there.” I made sense of these crowds [initially] as the “respectables,” the “non-respectables,” and the “near-respectables.” . . . Inside, such non-respectables might sit on the crates, but if a respectable came along and wanted to sit there, the lower status person would have to move. (pp. 225–226)

But this initial conceptualization changed with experience as Anderson (2003: 230) realized that the participants themselves used other terms to differentiate

social status: *winehead, hoodlum,* and*regular.* What did they mean by these terms? “The ‘regulars’ basically valued ‘decency.’ They associated decency with conventionality but also with ‘working for a living,’ or having a ‘visible means of support’” (p. 231). In this way, Anderson progressively refined his concept as he gained experience in the setting.

Howard S. Becker (1958) provides another excellent illustration of this iterative process of conceptualization in his study of medical students:

> When we first heard medical students apply the term “crock” to patients, we made an effort to learn precisely what they meant by it. We found, through interviewing students about cases both they and the observer had seen, that the term referred in a derogatory way to patients with many subjective symptoms but no discernible physical pathology. Subsequent observations indicated that this usage was a regular feature of student behavior and thus that we should attempt to incorporate this fact into our model of student-patient behavior. The derogatory character of the term suggested in particular that we investigate the reasons students disliked these patients. We found that this dislike was related to what we discovered to be the students’ perspective on medical school: the view that they were in school to get experience in recognizing and treating those common diseases most likely to be encountered in general practice. “Crocks,” presumably having no disease, could furnish no such experience. We were thus led to specify connections between the student-patient relationship and the student’s view of the purpose of his professional education. Questions concerning the genesis of this perspective led to discoveries about the organization of the student body and communication among students, phenomena which we had been assigning to another [segment of the larger theoretical model being developed]. Since “crocks” were also disliked because they gave the student no opportunity to assume medical responsibility, we were able to connect this aspect of the student-patient relationship with still another tentative model of the value system and hierarchical organization of the school, in which medical responsibility plays an important role. (p. 658)

In this excerpt, the researcher was first alerted to a concept by observations in the field, then refined his understanding of this concept by investigating its

meaning. By observing the concept's frequency of use, he came to realize its importance. Finally, he incorporated the concept into an explanatory model of student-patient relationships.

A well-designed chart, or **matrix**, can facilitate the coding and categorization process. Exhibit 10.2 shows an example of a coding form designed by Miles and Huberman (1994: 93–95) to represent the extent to which teachers and teachers' aides ("users") and administrators at a school gave evidence of various supporting conditions that indicated preparedness for a new reading program. The matrix condenses data into simple categories, reflects further analysis of the data to identify "degree" of support, and provides a multidimensional summary that will facilitate subsequent, more intensive analysis. Direct quotes still impart some of the flavor of the original text.

Exhibit 10.2 Example of Checklist Matrix

Presence of Supporting Conditions		
Condition	**For Users**	**For Administrators**
Commitment	*Strong*—"wanted to make it work."	*Weak* at building level. Prime movers in central office committed; others not.
Understanding	"*Basic*" ("felt I could do it, but I just wasn't sure how.") for teacher. *Absent* for aide ("didn't understand how we were going to get all this.")	*Absent* at building level and among staff. *Basic* for 2 prime movers ("got all the help we needed from developer.") *Absent* for other central office staff.
Materials	*Inadequate:* ordered late, puzzling ("different from anything I ever used"), discarded.	N.A.
Front-end training	"*Sketchy*" for teacher ("it all happened so quickly"); no demo class. *None* for aide ("totally unprepared. I had to learn along with the children.")	Prime movers in central office had training at developer site; none for others.
Skills	*Weak-adequate* for teacher. "*None*" for aide.	One prime mover (Robeson) skilled in substance; others unskilled.
Ongoing inservice	*None*, except for monthly committee meeting; no substitute funds.	*None*
Planning, coordination time	*None*: both users on other tasks during day; lab tightly scheduled, no free time.	*None*
Provisions for debugging	*None* systematized; spontaneous work done by users during summer.	*None*
School admin. support	*Adequate*	N.A.
Central admin. support	*Very strong* on part of prime movers.	Building admin. only acting on basis of central office commitment.
Relevant prior experience	*Strong* and useful in both cases: had done individualized instruction, worked with low achievers. But [the] aide [had] no diagnostic experience.	*Present* and useful in central office, esp. Robeson (specialist).

Source: Miles, Matthew B., and A. Michael Huberman. 1994. Qualitative data analysis, 2nd ed. Thousand Oaks, CA: Sage. Used with permission.

Matrix: A chart used to condense qualitative data into simple categories and provide a multidimensional summary that will facilitate subsequent, more intensive analysis.

Examining Relationships and Displaying Data

Examining relationships is the centerpiece of the analytic process because it allows the researcher to move from simple description of the people and settings to explanations of why things happened as they did with those people in that setting. A matrix can show how different concepts are related or, perhaps, what causes are linked with what effects.

In Exhibit 10.3, a matrix relates stakeholders' stake in a new program with the researcher's estimate of their attitude toward the program. Each cell of the matrix was to be filled in with a summary of an illustrative case study. In other matrix analyses, quotes might be included in the cells to represent the opinions of these different stakeholders, or the number of cases of each type might appear in the cells. The possibilities are almost endless. Keeping this approach in mind will generate many fruitful ideas for structuring a qualitative data analysis.

Exhibit 10.3 Coding Form for Relationships: Stakeholders' Stakes

	Estimate of Various Stakeholders' Inclination Toward the Program		
How high are the stakes for various primary stakeholders?	**Favorable**	**Neutral or Unknown**	**Antagonistic**
High			
Moderate			
Low			

Note: Construct illustrative case studies for each cell based on fieldwork.

Source: Patton, Michael Quinn. 2002. Qualitative research & evaluation methods, 3rd ed. Thousand Oaks, CA: Sage. Used with permission.

The simple relationships that are identified with a matrix like that shown in Exhibit 10.3 can be examined and then extended to create a more complex causal model. Such a model can represent the multiple relationships among the

important explanatory constructs. A great deal of analysis must precede the construction of such a model with careful attention to identification of important variables and the evidence that suggests connections between them. Exhibit 10.4 provides an example from a study of the implementation of a school program.

Research|Social Impact Link

Read more about examining relationships.

Authenticating Conclusions

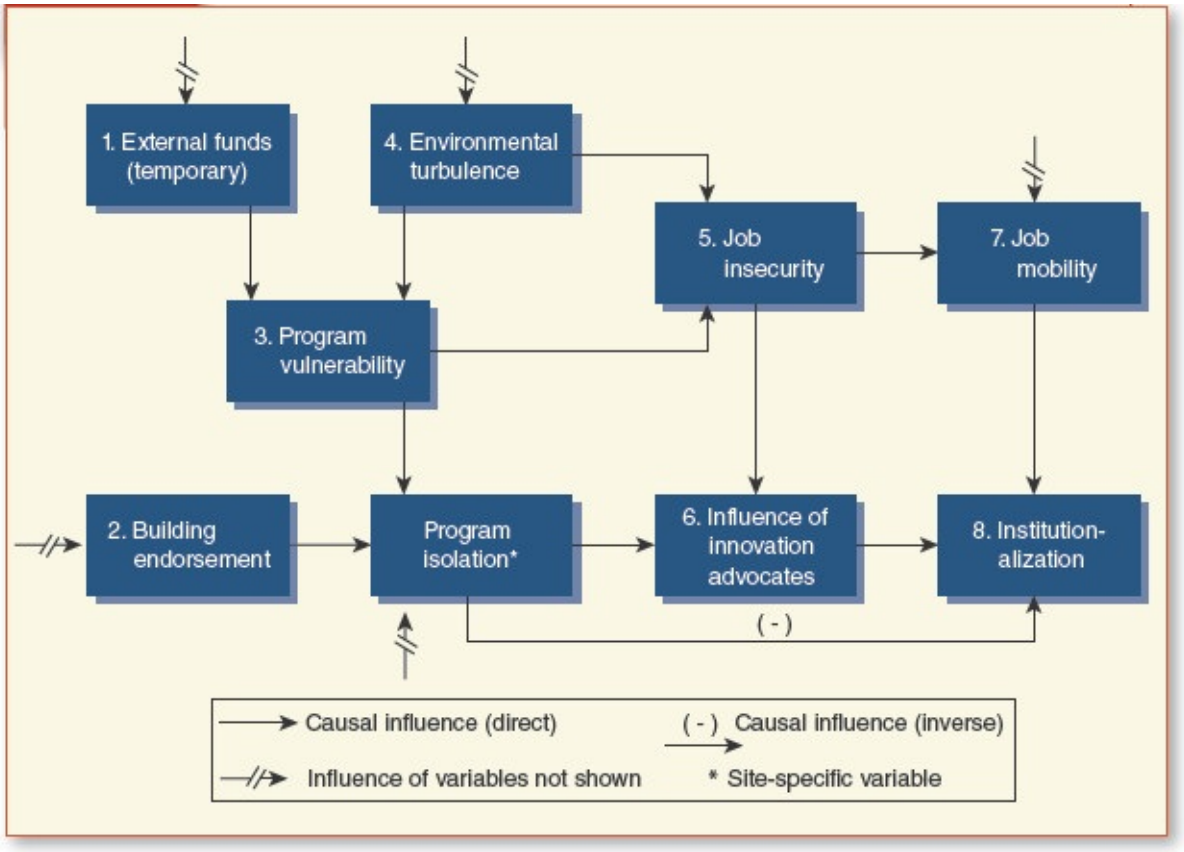
No set standards exist for evaluating the validity or *authenticity* of conclusions in a qualitative study, but the need to consider carefully the evidence and methods on which conclusions are based is just as great as with other types of research. Individual items of information can be assessed using at least three criteria (Becker 1958):

Journal Link

Read about a qualitative study that explores perceptions of masculinity. Are their conclusions authentic?

1. How credible was the informant? Were statements made by someone with whom the researcher had a relationship of trust or by someone the researcher had just met? Did the informant have reason to lie? If the statements do not seem to be trustworthy as indicators of actual events, can they at least be used to help understand the informant's perspective?
2. Were statements made in response to the researcher's questions, or were they spontaneous? Spontaneous statements are more likely to indicate what would have been said had the researcher not been present.
3. How does the presence or absence of the researcher or the researcher's informant influence the actions and statements of other group members? Reactivity to being observed can never be ruled out as a possible explanation for some directly observed social phenomenon. However, if the researcher carefully compares what the informant says goes on when the researcher is not present, what the researcher observes directly, and what other group members say about their normal practices, the extent of reactivity can be assessed to some extent.

Exhibit 10.4 Example of a Causal Network Model



Source: Miles, Matthew B., and A. Michael Huberman. 1994. Qualitative data analysis, 2nd ed. Thousand Oaks, CA: Sage. Used with permission.

A qualitative researcher's conclusions should also be judged by their ability to explain credibly some aspect of social life. Explanations should capture group members' tacit knowledge of the social processes that were observed, not just their verbal statements about these processes. **Tacit knowledge**—"the largely unarticulated, contextual understanding that is often manifested in nods, silences, humor, and naughty nuances"—is reflected in participants' actions as well as their words and in what they fail to state but nonetheless feel deeply and even take for granted (Altheide & Johnson 1994: 492–493). These features are evident in William F. Whyte's (1955) analysis of Cornerville social patterns:

> The corner-gang structure arises out of the habitual association of the members over a long period of time. The nuclei of most gangs can be traced back to early boyhood. . . . Home plays a very small role in the group

> activities of the corner boy . . . .
>
> The life of the corner boy proceeds along regular and narrowly circumscribed channels. . . . Out of [social interaction within the group] arises a system of mutual obligations which is fundamental to group cohesion. . . . The code of the corner boy requires him to help his friends when he can and to refrain from doing anything to harm them. When life in the group runs smoothly, the obligations binding members to one another are not explicitly recognized. (pp. 255–257)

Comparing conclusions from a qualitative research project to those other researchers obtained by conducting similar projects can also increase confidence in their authenticity. Miller's 1999 study of neighborhood police officers (NPOs) found striking parallels in the ways they defined their masculinity to processes reported in research about males in nursing and other traditionally female jobs (as cited in Bachman & Schutt 2007):

> In part, male NPOs construct an exaggerated masculinity so that they are not seen as feminine as they carry out the social-work functions of policing. Related to this is the almost defiant expression of heterosexuality, so that the men's sexual orientation can never truly be doubted even if their gender roles are contested. Male patrol officers' language—such as their use of terms like "pansy police" to connote neighborhood police officers—served to affirm their own heterosexuality. In addition, the male officers, but not the women, deliberately wove their heterosexual status into conversations, explicitly mentioning their female domestic partner or spouse and their children. This finding is consistent with research conducted in the occupational field. The studies reveal that men in female-dominated occupations, such as teachers, librarians, and pediatricians, over-reference their heterosexual status to ensure that others will not think they are gay. (p. 307)

Research That Matters

The Sexual Experiences Survey (SES) is used on many college campuses to assess the severity of sexual victimization, but researchers have found that it does not differentiate well between situations of unwanted sexual contact and attempted rape. Jenny Rinehart and Elizabeth Yeater (2011: 927) at the University of New Mexico designed a project to develop "a deeper qualitative understanding of the details of the event, as well as the context surrounding it."

As part of a larger study of dating experiences at a West Coast university, Rinehart and Yeater analyzed written narratives provided by 78 women who had given indicated some experience with sexual victimization on the SES. The authors and an undergraduate research assistant read each of the narratives and identified eight different themes and contexts, such as "relationship with the perpetrator." Next, they developed specific codes to make distinctions within each of the themes and contexts, such as "friend," "boss," or "stranger" within the "relationship" theme.

Here is an incident in one narrative that Rinehart and Yeater (2011: 934) coded as involving unwanted sexual contact with a friend:

> I went out on a date with a guy (he was 24) and we had a good time. He invited me into his apartment after to "hang out" for a little while longer. He tried pressuring me into kissing him at first, even though I didn't want to. Then he wrestled me (playfully to him, but annoyingly and unwanted to me). I repeatedly asked him to get off of me, and eventually he did. I kissed him once.

Their analysis of these narratives made it clear that incidents that received the same SES severity rating often differed considerably when the particulars were examined.

Source: Adapted from Rinehart, Jenny K., and Elizabeth A. Yeater. 2011. A qualitative analysis of sexual victimization narratives. *Violence Against Women* 17(7): 925–943.

Tacit knowledge: In field research, a credible sense of understanding of social processes that reflects the researcher's awareness of participants' actions, as well as their words, and of what they fail to state, feel deeply, and take for granted.

Reflexivity

Confidence in the conclusions from a field research study is also strengthened by an honest and informative account about how the researcher interacted with subjects in the field, what problems she encountered, and how these problems were or were not resolved. Such a "natural history" of the development of the evidence enables others to evaluate the findings. Such an account is important primarily because of the evolving and variable nature of field research: To an important extent, the researcher "makes up" the method in the context of a particular investigation rather than applying standard procedures that are specified before the investigation begins.

Barrie Thorne (1993) provides a good example of this final element of the analysis:

Encyclopedia Link

Read about the role of reflexivity in qualitative data analysis.

> Many of my observations concern the workings of gender categories in social life. For example, I trace the evocation of gender in the organization of everyday interactions, and the shift from boys and girls as loose aggregations to "the boys" and "the girls" as self-aware, gender-based groups. In writing about these processes, I discovered that different angles of vision lurk within seemingly simple choices of language. How, for example, should one describe a group of children? A phrase like "six girls and three boys were chasing by the tires" already assumes the relevance of gender. An alternative description of the same event—"nine fourth-graders were chasing by the tires"—emphasizes age and downplays gender. Although I found no tidy solutions, I have tried to be thoughtful about such choices. . . . After several months of observing at Oceanside, I realized that my field notes were peppered with the words "child" and "children," but that the children themselves rarely used the term. "What do they call themselves?" I badgered in an entry in my field notes. The answer it turned out, is that children use the same practices as adults. They refer to one

> another by using given names (“Sally,” “Jack”) or language specific to a given context (“that guy on first base”). They rarely have occasion to use age-generic terms. But when pressed to locate themselves in an age-based way, my informants used “kids” rather than “children.” (pp. 8–9)

Qualitative data analysts, more often than quantitative researchers, display real sensitivity to how a social situation or process is interpreted from a particular background and set of values and not simply based on the situation itself (Altheide & Johnson 1994). Researchers are only human, after all, and must rely on their own senses and process all information through their own minds. By reporting how and why they think they did what they did, they can help others determine whether, or how, the researchers’ perspectives influenced their conclusions.

Interactive Exercises

Qualitative Data Analysis

Anderson’s (2003) memoir about the Jelly’s Bar research illustrates the type of “tracks” that an ethnographer makes, as well as how the ethnographer can describe those tracks. Anderson acknowledges that his tracks began as a child:

> While growing up in the segregated black community of South Bend, from an early age, I was curious about the goings on in the neighborhood, but particularly streets, and more particularly, the corner taverns that my uncles and my dad would go to hang out and drink in. . . . Hence, my selection of Jelly’s as a field setting was a matter of my background, intuition, reason, and with a little bit of luck. (pp. 217–218)

After starting to observe at Jelly’s, Anderson’s (2003) “tracks” led to Herman:

> After spending a couple of weeks at Jelly’s, I met Herman and I felt that our meeting marked a big achievement. We would come to know each other well. . . . [He was] something of an informal leader at Jelly’s. . . . We were becoming friends. . . . He seemed to genuinely like me, and he was one

person I could feel comfortable with. (pp. 218–219)

Anderson's (2003) observations were shaped in part by Herman's perspective, but we also learn here that Anderson maintained some engagement with fellow students. This contact outside the bar helped to shape his analysis: "By relating my experiences to my fellow students, I began to develop a coherent perspective or a 'story' of the place which complemented the accounts that I had detailed in my accumulating field notes" (p. 220).

So, Anderson's analysis came in part from the way in which he "played his role" as a researcher and participant, not just from the setting itself.

What Are Some Alternatives in Qualitative Data Analysis?

The qualitative data analyst can choose from many interesting alternative approaches. Of course, the research question should determine the approach, but a researcher’s preferences will also inevitably play a role as well. The alternative approaches we present here (narrative analysis, conversation analysis, and grounded theory) will give you a good sense of the possibilities (Patton 2002).

Narrative Analysis

Narrative "displays the goals and intentions of human actors; it makes individuals, cultures, societies, and historical epochs comprehensible as wholes" (Richardson 1995: 200). **Narrative analysis** focuses on "the story itself" and seeks to preserve the integrity of personal biographies or a series of events that cannot adequately be understood in terms of their discrete elements (Riessman 2002: 218). The coding for a narrative analysis is typically of the narratives as a whole rather than of the different elements within them. The coding strategy revolves around reading the stories and classifying them into general patterns.

For example, Morrill and his colleagues (2000) read through 254 conflict narratives written by ninth graders (mentioned at the beginning of this chapter) and found four different types of stories:

1. Action tales, in which the author represents himself or herself and others as acting within the parameters of taken-for-granted assumptions about what is expected for particular roles among peers
2. Expressive tales, in which the author focuses on strong, negative emotional responses to someone who has wronged him or her
3. Moral tales, in which the author recounts explicit norms that shaped his or her behavior in the story and influenced the behavior of others
4. Rational tales, in which the author represents himself or herself as a rational decision maker navigating through the events of the story (p. 534)

Morrill et al. (2000: 534–535) also classified the stories along four stylistic dimensions: (1) plot structure (such as whether the story unfolds sequentially), (2) dramatic tension (how the central conflict is represented), (3) dramatic resolution (how the central conflict is resolved), and (4) predominant outcomes (how the story ends). Coding reliability was checked through a discussion by the two primary coders, who found that their classifications agreed for a large percentage of the stories.

The excerpt that begins this chapter exemplifies what Morrill et al. (2000: 536) termed an "action tale." Such tales

> unfold in matter-of-fact tones kindled by dramatic tensions that begin with a

> disruption of the quotidian order of everyday routines. A shove, a bump, a look . . . triggers a response. . . . Authors of action tales typically organize their plots as linear streams of events as they move briskly through the story's scenes. . . . This story's dramatic tension finally resolves through physical fighting, but . . . only after an attempted conciliation. (p. 356)

Audio Link

Listen to learn more about narrative analysis.

You can contrast that "action tale," with the following narrative, which Morrill et al. (2000: 545–546) classify as a "moral tale," in which the student authors "explicitly tell about their moral reasoning, often referring to how normative commitments shape their decision making":

> I . . . got into a fight because I wasn't allowed into the basketball game. I was being harassed by the captains that wouldn't pick me and also many of the players. The same type of things had happened almost every day where they called me bad words so I decided to teach the ring leader a lesson. I've never been in a fight before but I realized that sometimes you have to make a stand against the people that constantly hurt you, especially emotionally. I hit him in the face a couple of times and I got respect I finally deserved. (pp. 545–546)

Morrill et al. (2000: 553) summarize their classification of the youth narratives in a simple table that highlights the frequency of each type of narrative and the characteristics associated with each of them (Exhibit 10.5). How does such an analysis contribute to our understanding of youth violence? Morrill et al. first emphasize that their narratives "suggest that consciousness of conflict among youths—like that among adults—is not a singular entity but comprises a rich and diverse range of perspectives" (p. 551).

Exhibit 10.5 Summary Comparison of Youth Narratives*

Representation of	Action Tales (N = 144)	Moral Tales (N = 51)	Expressive Tales (N = 35)	Rational Tales (N = 24)
Bases of everyday conflict	disruption of everyday routines & expectations	normative violation	emotional provocation	goal obstruction
Decision making	intuitive	principled stand	sensual	calculative choice
Conflict handling	confrontational	ritualistic	cathartic	deliberative
Physical violence†	in 44% (*N* = 67)	in 27% (*N* = 16)	in 49% (*N* = 20)	in 29% (*N* = 7)
Adults in youth conflict control	invisible or background	sources of rules	agents of repression	institutions of social control

*Total N = 254.

†Percentages based on the number of stories in each category.

Source: Morrill, Calvin, Christine Yalda, Madeleine Adelman, Michael Musheno, and Cindy Bejarano. 2000. Telling tales in school: Youth culture and conflict narratives. Law & Society Review 34: 521–565. Reprinted with permission of Blackwell Publishing Ltd.

Theorizing inductively, Morrill et al. (2000: 553–554) then attempt to explain why action tales were much more common than were the more adult-oriented normative, rational, or emotionally expressive tales. They say that one possibility is to be found in Carol Gilligan's theory of moral development, which suggests that younger students are likely to limit themselves to the simpler action tales that "concentrate on taken-for-granted assumptions of their peer and wider cultures, rather than on more self-consciously reflective interpretation and evaluation" (pp. 553–554). More generally, Morrill et al. argue, "We can begin to think of the building blocks of cultures as different narrative styles in which various aspects of reality are accentuated, constituted, or challenged, just as others are deemphasized or silenced" (p. 556).

In this way, Morrill et al.'s narrative analysis allowed an understanding of youth

conflict to emerge from the youths' own stories while informing our understanding of broader social theories and processes.

Narrative analysis: A form of qualitative analysis in which the analyst focuses on how respondents impose order on the flow of experience in their lives and so make sense of events and actions in which they have participated.

Conversation Analysis

Conversation analysis is a specific qualitative method for analyzing ordinary conversation. Unlike narrative analysis, conversation analysis focuses on the sequence and details of conversational interaction rather than on the "stories" that people are telling. Like ethnomethodology, from which it developed, conversation analysis focuses on how reality is constructed rather than on what it "is."

Three premises guide conversation analysis (Gubrium & Holstein 2000):

1. Interaction is sequentially organized, and talk can be analyzed in terms of the process of social interaction rather than in terms of motives or social status.
2. Talk, as a process of social interaction, is contextually oriented—it both is shaped by interaction and creates the social context of that interaction.
3. These processes are involved in all social interaction, so no interactive details are irrelevant to understanding it. (p. 492)

Consider these premises as you read the following dialogue between British researcher Ann Phoenix (2003) and a boy she called "Thomas" in her study of notions of masculinity, bullying, and academic performance among 11- to 14-year-old boys in 12 London schools:

Thomas: It's your attitude, but some people are bullied for no reason whatsoever just because other people are jealous of them. . . .

Q: How do they get bullied?

Thomas: There's a boy in our year called James, and he's really clever and he's basically got no friends, and that's really sad because . . . he gets top marks in every test and everyone hates him. I mean, I like him. . . . (p. 235)

Phoenix (2003) notes that here,

> Thomas dealt with the dilemma that arose from attempting to present himself as both a boy and sympathetic to school achievement. He . . .

> distanced himself from . . . being one of those who bullies a boy just because they are jealous of his academic attainments . . . constructed for himself the position of being kind and morally responsible. (p. 235)

Note that Thomas was a boy talking to a woman. Do you imagine that his talk would have been quite different if his conversation had been with other boys?

An example of the very detailed data recorded in a formal conversation analysis appears in Exhibit 10.6. It is from David R. Gibson's (2005: 1566) study of the effects of superior–subordinate and friendship interaction on the transitions that occur during conversation—in this case, in meetings of managers. Every type of "participation-shift" (P-shift) is recorded and distinguished from every other type. Some shifts involve "turn claiming," in which one person (X) begins to talk after the first person (A) has addressed the group as a whole (0), without being prompted by the first speaker. Some shifts involve "turn receiving," in which the first person (A) addresses the second (B), who then responds. In "turn usurping," by contrast, the second person (X) speaks after the first person (A) has addressed a comment to a third person (B), who is thus prevented from responding. Examining this type of data can help us to see how authority is maintained or challenged in social groups.

Exhibit 10.6 Inventory of P-Shifts With Examples

P-Shift	Example
Turn claiming:	
AO-XA.....	John talks to the group, then Frank talks to John.
AO-XO.....	John talks to the group, then Frank talks to the group.
AO-XY.....	John talks to the group, then Frank talks to Mary.
Turn receiving:	
AB-BA.....	John talks to Mary, then Mary replies.
AB-BO.....	John talks to Mary, then Mary talks to the group.
AB-BY.....	John talks to Mary, then Mary talks to Irene.
Turn usurping:	
AE-XA.....	John talks to Mary, then Frank talks to John.
AB-XB.....	John talks to Mary, then Frank talks to Mary.
AB-XO.....	John talks to Mary, then Frank talks to the group.
AB-XY.....	John talks to Mary, then Frank talks to Irene.

Note: The initial speaker is denoted A and the initial target B, unless the group is addressed (or the target was ambiguous), in which case the target is O. Then, the P-shift is summarized in the form (speaker1) (target1)-(speaker2) (target2), with A or B appearing after the hyphen only if the initial speaker or target serves in one of these two positions in the second turn. When the speaker in the second turn is someone other than A or B, X is used, and when the target in the second turn is someone other than A, B, or the group O, Y is used.

Source: Gibson, David R. 2005. Taking turns and talking ties: Networks and conversational interaction. American Journal of Sociology 110(6): 1561–1597. Copyright © 2005 The University of Chicago. Reprinted with permission from the University of Chicago Press.

Grounded Theory

Theory development occurs continually in qualitative data analysis (Coffey & Atkinson 1996: 23). The goal of many qualitative researchers is to create **grounded theory**—that is, to build up inductively a systematic theory that is "grounded" in, or based on, the observations. The observations are summarized into conceptual categories, which are tested directly in the research setting with more observations. Over time, as the conceptual categories are refined and linked, a theory evolves (Glaser & Strauss 1967; Huberman & Miles 1994: 436).

As observation, interviewing, and reflection continue, researchers refine their definitions of problems and concepts and select indicators. They can then check the frequency and distribution of phenomena: How many people made a particular type of comment? How often did social interaction lead to arguments? Social system models may then be developed, which specify the relationships among different phenomena. These models are modified as researchers gain experience in the setting. For the final analysis, the researchers check their models carefully against their notes and make a concerted attempt to discover negative evidence that might suggest the model is incorrect.

Grounded theory: Systematic theory developed inductively, based on observations that are summarized into conceptual categories, reevaluated in the research setting, and gradually refined and linked to other conceptual categories.

Careers and Research

Laurel Person Mecca, MA, Assistant Director and Senior Research Specialist, Qualitative Data Analysis Program

Source: Laurel Person Mecca

Laurel Person Meca was uncertain of the exact career she wanted to pursue during her graduate studies at the Louisiana State University. Then she happened upon the University Center for Social & Urban Research (UCSUR). It's hard to imagine a better place to launch a research career involving qualitative data analysis. Since 2005, the center has provided services and consultation to investigators in qualitative data analysis. Mecca used UCSUR to recruit participants for her own research and then made it clear to staff that she would love to work there after finishing her degree. Fourteen years later, she enjoys her work there more than ever.

One of the greatest rewards Mecca has found in her work is the excitement of discovering the unexpected when her preconceived notions about what research participants will tell her turn out to be incorrect. She also finds that her interactions with research participants provide a unique view into peoples' lives, thus providing insights in her own life and a richer understanding of the human condition. And in addition to these personal benefits, Mecca has the satisfaction of seeing societal benefits from the projects she consults on: improving technologies designed to enhance independent living about elderly and disabled persons; exploring the barriers to participation in the Supplemental Nutrition Assistance Program (SNAP); evaluating a program to improve parent–adolescent communication about sexual behaviors to reduce sexually transmitted diseases (STDs) and unintended teen pregnancies.

Mecca has some very sound advice for students interested in careers involving doing research or using research results:

> Gain on-the-job experience while in college, even if it is an unpaid internship. Find researchers who are conducting studies that interest you, and inquire about working for them. Even if they are not posting an available position, they may bring you on board. Persistence pays off! You are much more likely to be selected for a position if you demonstrate a genuine interest in the work and if you continue to show your enthusiasm by following up.
>
> Definitely check out the National Science Foundation's (NSF) Research Experience for Undergraduates (REU) program. Though most of these internships are in the "hard" sciences, there are plenty of openings in social sciences disciplines. These internships include a stipend, and oftentimes, assistance with travel and housing. They are wonderful opportunities to work directly on a research project, and may provide the additional benefit of a conference presentation and/or publication.

Visual Sociology

The analysis of the “text” of social life, then, can be conducted in a variety of ways. But words are not the only form of qualitative data. For about 150 years, people have been recording the social world with photography. This creates the possibility of “observing” the social world through photographs and films and of interpreting the resulting images as a text. Visual sociology is a method both to learn how others “see” the social world and to create images of it for further study. As with written text, however, the visual sociologist must be sensitive to the way in which a photograph or film “constructs” the reality that it depicts.

Audio Link

Listen to how visual sociology is used.

An analysis by Eric Margolis (2004) of photographic representations of American Indian boarding schools gives you an idea of the value of analysis of photographs. On the left is a picture taken in 1886 of Chiricahua Apaches who had just arrived at the Carlisle Indian School in Carlisle, Pennsylvania (Exhibit 10.7). The school was run by Captain Richard Pratt, who, like many Americans in that period, felt that tribal societies were communistic, indolent, dirty, and ignorant, whereas Western civilization was industrious and individualistic. So Pratt set out to acculturate American Indians to the dominant culture. The second picture shows the result: the same group of Apaches looking like European, not Native, Americans, dressed in “standard” (per the dominant culture) uniforms with standard haircuts and with more standard posture.

Exhibit 10.7 Pictures of Chiricahua Apache Children Before and After Starting Carlisle Indian School, Carlisle, Pennsylvania, 1886

Source: Choate, J. N. (John N.); Western History/Genealogy Dept., Denver Public Library; and Beinecke Rare Book & Manuscript Library, Yale University.

Many other pictures display the same type of transformation. Are these pictures each “worth a thousand words”? They capture the ideology of the school management, but we can be less certain that they document accurately the “before and after” status of the students. Pratt “consciously used photography to represent the boarding school mission as successful” (Margolis 2004: 79). Although he clearly tried to ensure a high degree of conformity, there were accusations that the contrasting images were exaggerated to overemphasize the change (Margolis 2004: 78). In these photographs, reality was being constructed, not just depicted.

With the widespread use of cell phone cameras and video recorders, visual sociology will certainly become an increasingly important aspect of qualitative analyses of social settings and the people in them. The result will be richer descriptions of the social world, but remember Darren Newbury’s (2005) reminder to readers of his journal, *Visual Studies*: “Images cannot be simply taken of the world, but have to be made within it” (p. 1).

Why Are Mixed Methods Helpful?

Different methods have different strengths and weaknesses. Using *mixed methods* can reinforce each other, create a greater depth of understanding, reveal or correct errors in other methods, and fill in the steps in complex social processes.

Sometimes new methods are introduced to replicate and strengthen existing research findings. Susan McCarter (2009) extended prior research on juvenile justice processing with an integrated mixed-methods investigation of case processing and participant orientations in Virginia.

The large quantitative data set McCarter (2009) used in her research was secondary data collected on 2,233 African American and Caucasian males in Virginia's juvenile justice system, covering

> juveniles' previous felonies; previous misdemeanors; previous violations of probation/parole; previous status offenses; recent criminal charges, intake action on those charges, pre-disposition(s) of those charges, court disposition(s) of those charges; and demographics such as sex, race, data of birth, [Court Service Unit] CSU, and geotype (urban, suburban, rural). For a subset of these cases, data also included information from the youth's social history, which required judicial request. (p. 535)

Qualitative data, on the other hand, were obtained from 24 in-depth interviews with juvenile judges, the commonwealth's attorneys, defense attorneys, police officers, juveniles, and their families (McCarter 2009):

> The juvenile justice personnel were from six Court Service Units across the state, including two urban, two suburban, two rural, two from Region I, two from Region II, and two from Region III. . . . Participants from each CSU were chosen to provide maximum diversity in perspectives and experiences, and thus varied by race, sex, and age; and the justice personnel also varied in length of employment, educational discipline and educational attainment. (p. 536). The youth and their families were all selected from one Court

> Service Unit (CSU) located in an urban geotype with a population of approximately 250,000 (p. 536). The sample of youth and their family members was comprised of all male juveniles, five mothers and one father. Four of the six families were African American and two were Caucasian. (p. 540)

The in-depth interviews included both open- and closed-ended questions. The open-ended responses are coded into categories that distinguished how participants perceived the role of race in the juvenile justice system (McCarter 2009: 536). A direct connection with the quantitative findings.

In the interviews themselves,

> respondents were read the quantitative findings from this study and then asked whether or not their experiences and/or perceptions of the juvenile justice system were congruent with the findings. They were also asked how commonly they believed instances of racial or ethnic bias occurred in Virginia. (McCarter 2009: 540)

The responses to this qualitative question supported the quantitative finding that race mattered:

> Juvenile justice professionals as well as youth and their families cited racial bias by individual decision-makers and by the overall system, and noted that this bias was most likely to occur by the police during the Alleged Act or Informal Handling stages. However, although race was considered a factor, when compared to other factors, professionals did not think race played a dominant role in affecting a youth's treatment within the juvenile justice system. . . . Eighteen of the juvenile justice professionals stated that they felt a disparity [between processing of African American and white juveniles] existed, four did not feel that a disparity existed, and two indicated that they did not know. (McCarter 2009: 540)

In this way, the qualitative and quantitative findings were integrated, and the study's key conclusion about race-based treatment was strengthened (McCarter

2009: 542).

Mixed methods can also deepen understanding of a phenomenon. After a devastating earthquake in İzmit, Turkey, on August 17, 1999, killed 19,000 people, Elif Kale-Lostuvali (2007) conducted research using a combination of qualitative methodologies—including participant observation and intensive interviewing—to study citizen–state encounters in the region.

One important concept that emerged from Kale-Lostuvali's observations and interviews was a distinction locals made between a *magdur* (sufferer) and a *depremzade* (son of the earthquake). This was a critical distinction because a *magdur* was seen as deserving of government assistance, whereas a *depremzade* was considered to be taking advantage of the situation for personal gain. Kale-Lostuvali (2007) drew on both interviews and participant observation to develop an understanding of this complex concept:

Video Link

Watch a video on understanding and developing a mixed methods approach.

> A prominent narrative frequently repeated in the disaster area elaborated the contrast between *magdur* (sufferer, that is, the truly needy) and *depremzade* (sons of the earthquake). The *magdur* (sufferers) were the deserving recipients of the aid that was being distributed. However, they (1) were in great pain and could not pursue what they needed; or (2) were proud and could not speak of their need; or (3) were humble, always grateful for the little they got, and were certainly not after material gains; or (4) were characterized by a combination of the preceding. And because of these characteristics, they had not been receiving their rightful share of the aid and resources. In contrast, *depremzade* (sons of the earthquake) were people who took advantage of the situation. (p. 755)

Similarly, the qualitative research by Spencer Moore and his colleagues (2004) on the social response to Hurricane Floyd combined data from focus groups and from participant observation with the workers. Reports of heroic acts by rescuers, innumerable accounts of "neighbors helping neighbors," and the

comments of Health Works After the Flood (task force) participants suggest that residents, stranded motorists, relief workers, and rescuers worked and came together in remarkable ways during the relief and response phases of the disaster:

> Like people get along bette.r . . . they can talk to each other. People who hadn't talked before, they talk now, a lot closer. That goes, not only for the neighborhood, job-wise, organization-wise, and all that. . . . [Our] union sent some stuff for some of the families that were flooded out. (Focus Group #4) (pp. 210–211)

Mixing methods can help offset the intrinsic weaknesses of each technique. For example, Renee Anspach (1991) wondered about the use of standard surveys to study the effectiveness of mental health systems. So instead of drawing a large sample and asking a set of closed-ended questions, Anspach used snowball sampling techniques to select some administrators, case managers, clients, and family members in four community mental health systems, and then asked these respondents a series of open-ended questions. When asked whether their programs were effective, the interviewees were likely to respond Yes, but their comments in response to other questions pointed to many program failings. Anspach concluded that the respondents simply wanted the interviewer (and others) to believe in the program's effectiveness, for several reasons: Administrators wanted to maintain funding and employee morale; case managers wanted to ensure cooperation by talking up the program with clients and their families; and case managers also preferred to deflect blame for problems to clients, families, or system constraints.

Mixed methods can help us understand complex issues such as violence against women, "a multifaceted phenomenon, occurring within a social context that is influenced by gender norms, interpersonal relationships, and sexual scripts," in which, as Maria Testa, Jennifer Livingston, and Carol VanZile-Tamsen (2011) report, "understanding of these experiences of violence is dependent on the subjective meaning for the woman and cannot easily be reduced to a checklist" (p. 237).

So Testa and her colleagues (2011) supplemented their quantitative study of violence against women with a qualitative component. Victims' responses to

structured survey questions showed a quantitative association between alcohol use and rape victimization. Such an association has often been interpreted as suggesting "impaired judgment" about consent by intoxicated victims. But Testa et al. (2011) found that rape usually occurred after excessive drinking when the women were truly incapacitated, and therefore could neither resist nor even be fully aware of what was happening. Testa and her colleagues concluded that the prevalence of this type of "incapacitated rape" required a new approach to the problem of violence against women (2011: 242):

> Qualitative analysis of our data has resulted in numerous "a-ha" types of insights that would not have been possible had we relied solely on quantitative data analysis (e.g., identification of incapacitated rape and sexual precedence, heterogeneity in the way that sexual assaults arise) and also helped us to understand puzzling quantitative observations. . . . These insights, in turn, led to testable, quantitative hypotheses that supported our qualitative findings, lending rigor and convergence to the process. (p. 245)

Even official documents (maybe especially such documents) can themselves be scrutinized with other methods, revealing what's really happening. Consider the court records of juveniles accused of illegal acts, which document the critical decisions to arrest, to convict, or to release (Dannefer & Schutt 1982). Research based on such records is only as good as the records themselves. As indicated in Exhibit 10.8, Carolyn Needleman's participant observation study of probation officers in two New York juvenile court intake units (1981) found that what researchers believe they are measuring with official records differs markedly from what probation officers mean by those records.

Researchers assume that sending a juvenile case to court indicates a more severe outcome than retaining a case in the intake unit, but probation officers often diverted cases from court because they thought the courts would be too lenient. Researchers assume that probation officers evaluate juveniles as individuals, but probation officers often based their decisions on juveniles' current social situation (e.g., whether they were living in a stable home), without learning anything about the individual juvenile. Perhaps most troubling, Needleman (1981) found that probation officers often decided how to handle cases first and then created an official record that appeared to justify their decisions.

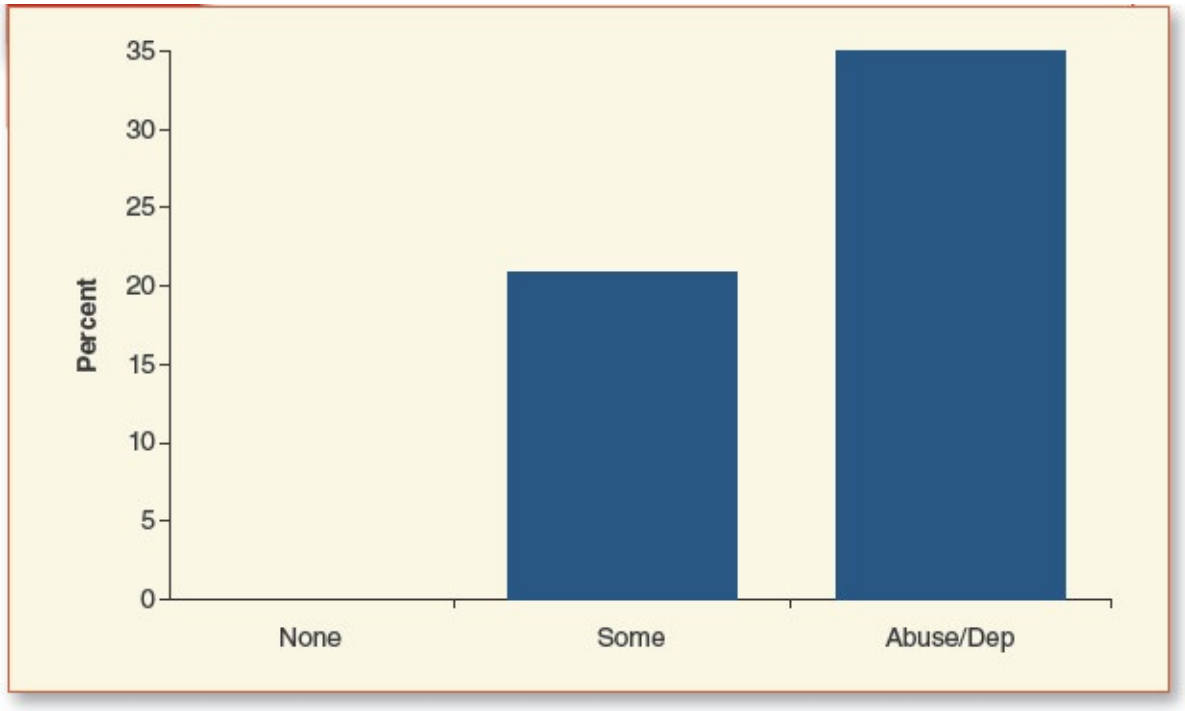
Exhibit 10.8 Researchers' and Juvenile Court Workers' Discrepant Assumptions

Researcher Assumptions	Probation Officer Assumptions
• Being sent to court is a harsher sanction than diversion from court. • Screening involves judgments about individual juveniles. • Official records accurately capture case facts.	• Being sent to court often results in more lenient and less effective treatment. • Screening should center on the juvenile's social situation. • Records can be manipulated to achieve the desired outcome.

Source: Needleman, Carolyn. 1981. Discrepant assumptions in empirical research: The case of juvenile court screening. Social Problems 28 (February): 248–256.

Different methods of research can also fill in different steps on a social process, better explaining overall outcomes. In Russ Schutt's study of homelessness and mental illness, he found a quantitative association between lifetime substance abuse—a diagnosis recorded on a numerical scale that was made on the basis of an interview with a clinician—and housing loss (Exhibit 10.9) (Schutt 2011: 135).

Exhibit 10.9 Substance Abuse and Housing Loss in Group Homes



Source: Schutt, Russell K., with Stephen M. Goldfinger. 2011. Homelessness, housing and mental illness. Cambridge, MA: Harvard University Press, 135 Copyright © 2011 by the President and Fellows of Harvard College. Used with permission.

Ethnographic notes recorded in the same group homes help explain the substance abuse–housing loss association (Schutt 2011):

> The time has come where he has to decide once and for all to drink or not. . . . Tom has been feeling "pinned to the bed" in the morning. He has enjoyed getting high with Sammy and Ben, although the next day is always bad. . . . Since he came back from the hospital Lisandro has been acting like he is taunting them to throw him out by not complying with rules and continuing to drink. . . . (pp. 131, 133)

The analysis of the quantitative data reveals *what* happened, and Schutt's analysis of the ethnographic data helps to understand *why*.

Finally, Dan Chambliss and Chris Takacs (2014), in a 5-year longitudinal study

of students' development of writing skills in college, used a combination of content analysis, surveys, and in-depth panel interviews to measure and understand how—and if—students actually improved their writing during their college careers. More than 1,000 papers, running from the final year of high school all the way through college, were assembled; they were "blind" graded by outside evaluators. Overall, students showed noticeable improvement during the first 3 years.

Analysis of quantified results on senior surveys then showed that the students who improved the most were aware of that improvement, and in the interviews, those students credited their improvement partly to one-on-one meetings—even a single meeting with a professor who cared about them and their work. A mixed-method study, then, was able to uncover the extent of students' learning, students' own ability to assess their learning, and the means by which the learning occurred—providing a well-rounded understanding of an important phenomenon.

Researcher Interview Link

Watch a researcher describe immigration patterns using comparative research.

Research in the News

What's in a Message?

In response to the large number of military suicides, Attivio Inc. and military suicide experts are creating a qualitative coding scheme and searching through social media posts. Facebook and Twitter posts hold valuable information about a person's well-being; researchers are creating a way to analyze millions of posts. The program is called the Durkheim Project and aims to prevent suicides by systematically identifying at-risk soldiers or veterans.

1. What are the advantages and disadvantages that you see in using Facebook and Twitter posts to determine a person's well-being?
2. How do you think that analyzing online social interaction would compare as a research method with analyzing social interaction in person, through interviews or observations? Which approach would be preferable in identifying at-risk soldiers or veterans?

News Source: Weintraub, Karen. 2013. Monitoring social media to cut the military suicide rate. *New York Times*, July 22:B5.

How Can Computers Assist Qualitative Data Analysis?

Computer-assisted qualitative data analysis can dramatically accelerate the techniques used traditionally to analyze such text as notes, documents, or interview transcripts: preparation, coding, analysis, and reporting (Coffey & Atkinson 1996; Richards & Richards 1994). Two of the most popular programs can illustrate these steps: HyperRESEARCH and QSR NVivo. (You can link to a trial copy of HyperRESEARCH and tutorials about it on the book's Study Site at edge.sagepub.com/chamblissmssw5e.)

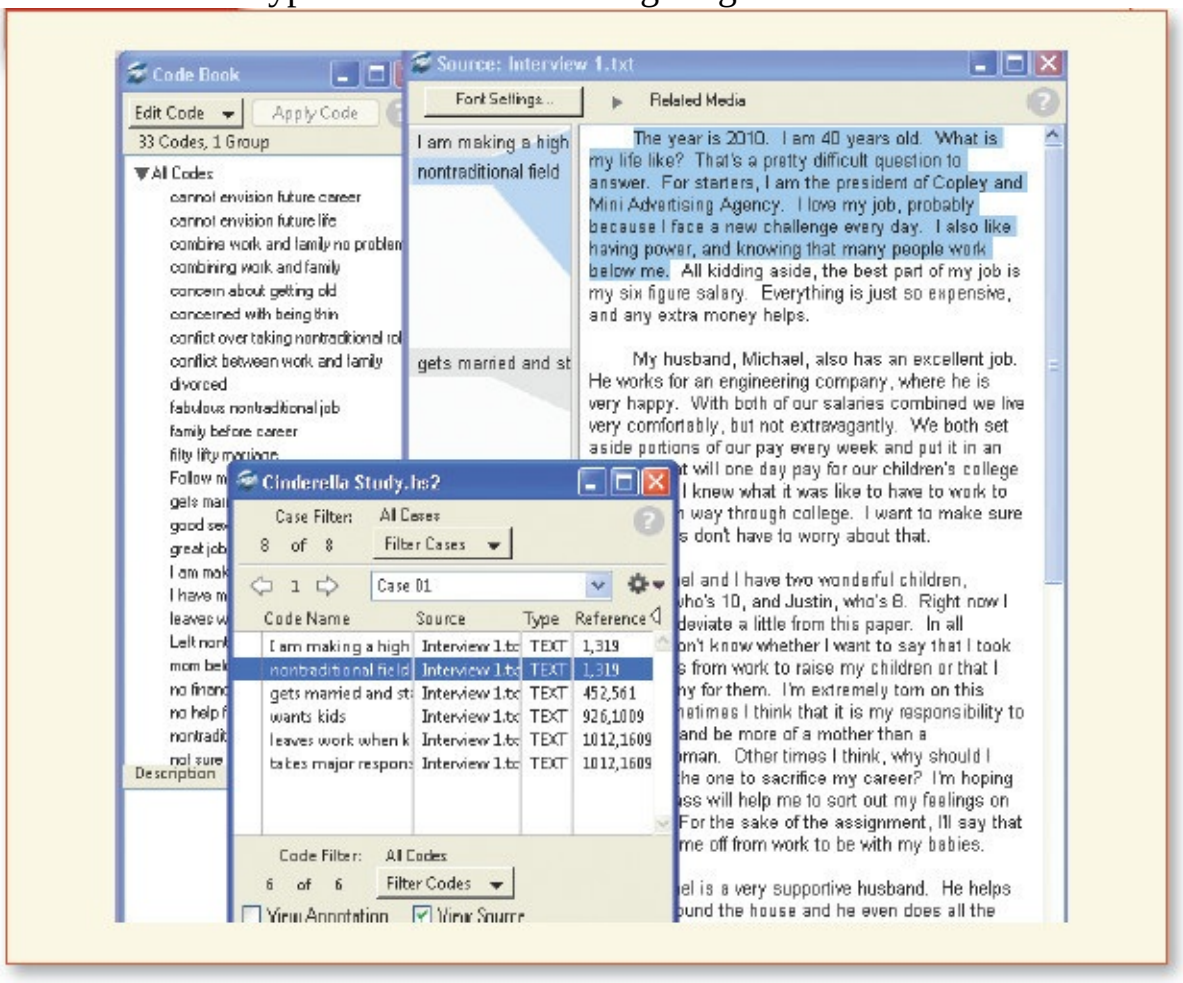
Text preparation begins with typing or scanning text in a word processor or, with NVivo, directly into the program's rich text editor. NVivo will create or import a rich text file (*.rtf). HyperRESEARCH requires that your text be saved as a text file (as "ASCII" in most word processors, or *.txt) before you transfer it into the analysis program. HyperRESEARCH expects your text data to be stored in separate files corresponding to each unique case, such as an interview with one subject.

Coding the text involves categorizing particular text segments. This is the foundation of much qualitative analysis. Either program allows you to assign a code to any segment of text (in NVivo, you drag through the characters to select them; in HyperRESEARCH, you click on the first and last words to select text). You can either make up codes as you go through a document or assign codes that you have already developed to text segments. Exhibits 10.12a and 10.12b show the screens that appear in the two programs at the coding stage, when a particular text segment is being labeled. You can also have the programs "autocode" text by identifying a word or phrase that should always receive the same code, or, in NVivo, by coding each section identified by the style of the rich text document—for example, each question or speaker. (Of course, you should check carefully the results of autocoding.) Both programs also let you examine the coded text "in context"—embedded in its place in the original document.

In qualitative data analysis, coding is not a one-time-only or one-code-only procedure. Both HyperRESEARCH and NVivo allow you to be inductive and holistic in your coding: You can revise codes as you go along, assign multiple

codes to text segments, and link your own comments (“memos”) to text segments. In NVivo, you can work “live” with the coded text to alter coding or create new, more subtle categories. You can also place hyperlinks to other documents in the project or any multimedia files outside it.

Exhibit 10.10a HyperRESEARCH Coding Stage


Code Book
Edit Code
Apply Code
33 Codes, 1 Group
All Codes
cannot envision future career
cannot envision future life
combining work and family
concern about getting old
conflict between work and family
divorced
fabulous nontraditional job
family before career
Description
Source: Interview 1.txt
Font Settings...
Related Media
I am making a high
nontraditional field
gets married and st
The year is 2010. I am 40 years old. What is my life like? That's a pretty difficult question to answer. For starters, I am the president of Copley and Mini Advertising Agency. I love my job, probably because I face a new challenge every day. I also like having power, and knowing that many people work below me. All kidding aside, the best part of my job is my six figure salary. Everything is just so expensive, and any extra money helps.
My husband, Michael, also has an excellent job. He works for an engineering company, where he is very happy. With both of our salaries combined we live very comfortably, but not extravagantly. We both set aside portions of our pay every week and put it in an
Cinderella Study.hs2
Case Filter: All Cases
8 of 8
Filter Cases
Case 01
Code Name
Source
Type
Reference
I am making a high
Interview 1.tc
TEXT
1,319
nontraditional field
Interview 1.tc
TEXT
1,319
gets married and st
Interview 1.tc
TEXT
452,561
wants kids
Interview 1.tc
TEXT
926,1009
leaves work when k
Interview 1.tc
TEXT
1012,1609
takes major respons
Interview 1.tc
TEXT
1012,1609
Code Filter: All Codes
6 of 6
Filter Codes
View Annotation
View Source

Analysis focuses on reviewing cases or text segments with similar codes and examining relationships among different codes. You may decide to combine codes into larger concepts. You may specify additional codes to capture more fully the variation among cases. You can test hypotheses about relationships among codes. NVivo allows development of an indexing system to facilitate thinking about the relationships among concepts and the overarching structure of these relationships. It also allows you to draw more free-form models (Exhibit 10.11). In HyperRESEARCH, you can specify combinations of codes that identify cases that you want to examine.

Reports from both programs can include text to illustrate the cases, codes, and relationships that you specify. You can also generate counts of code frequencies and then import these counts into a statistical program for quantitative analysis. However, the many types of analyses and reports that can be developed with qualitative analysis software do not lessen the need for a careful evaluation of the quality of the data on which conclusions are based.

In practice, using these programs is not always as time-saving as it may first appear (Bachman & Schutt 2007: 319). Scott Decker and Barrik Van Winkle (1996: 53–54) described the difficulty they faced in using a computer program to identify instances of “drug sales”:

> The software we used is essentially a text retrieval package. . . . One of the dilemmas faced in the use of such software is whether to employ a coding scheme within the interviews or simply to leave them as unmarked text. We chose the first alternative, embedding conceptual tags at the appropriate points in the text. An example illustrates this process. One of the activities we were concerned with was drug sales. Our first chore (after a thorough reading of all the transcripts) was to use the software to “isolate” all of the transcript sections dealing with drug sales. One way to do this would be to search the transcripts for every instance in which the word “drugs” was used. However, such a strategy would have the disadvantages of providing information of too general a character while often missing important statements about drugs. Searching on the word “drugs” would have produced a file including every time the word was used, whether it was in reference to drug sales, drug use, or drug availability, clearly more information than we were interested [in]. However, such a search would have failed to find all of the slang used to refer to drugs (“boy” for heroin, “Casper” for crack cocaine) as well as the more common descriptions of drugs, especially rock or crack cocaine.

Exhibit 10.10 b NVivo Coding Stage

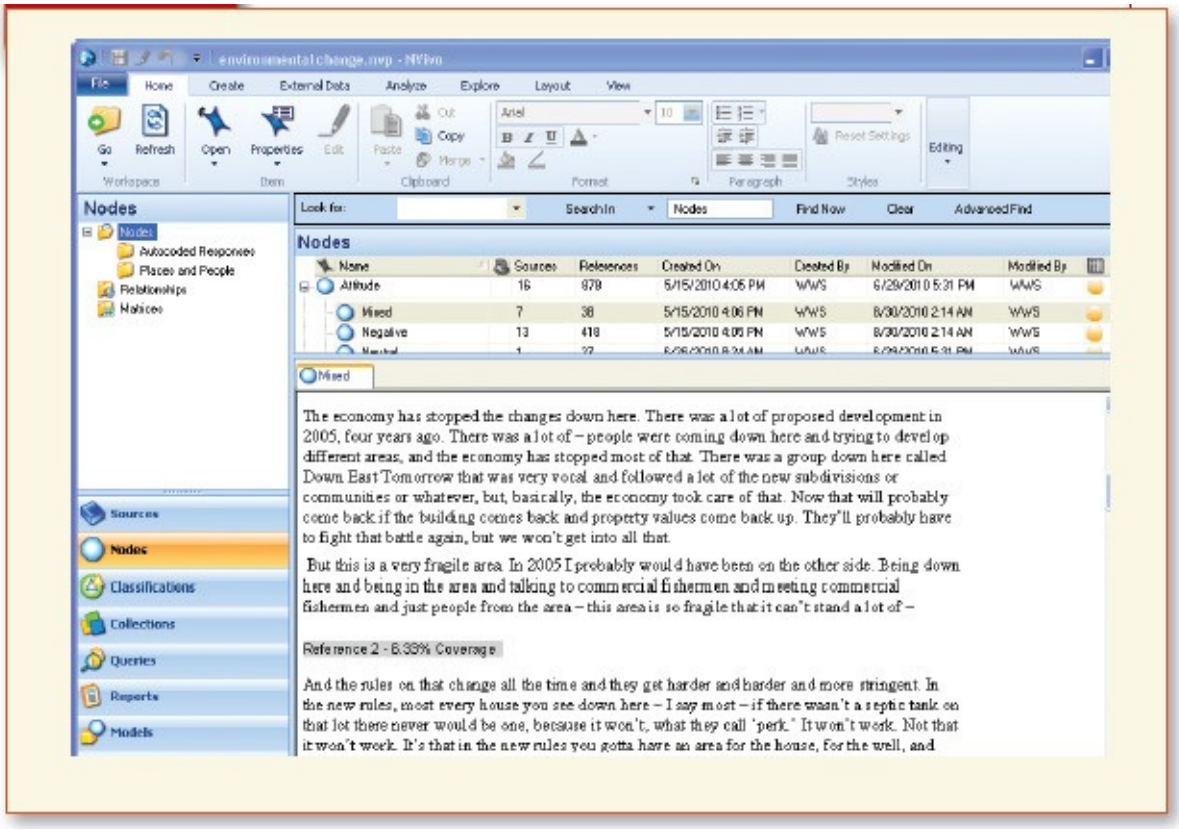
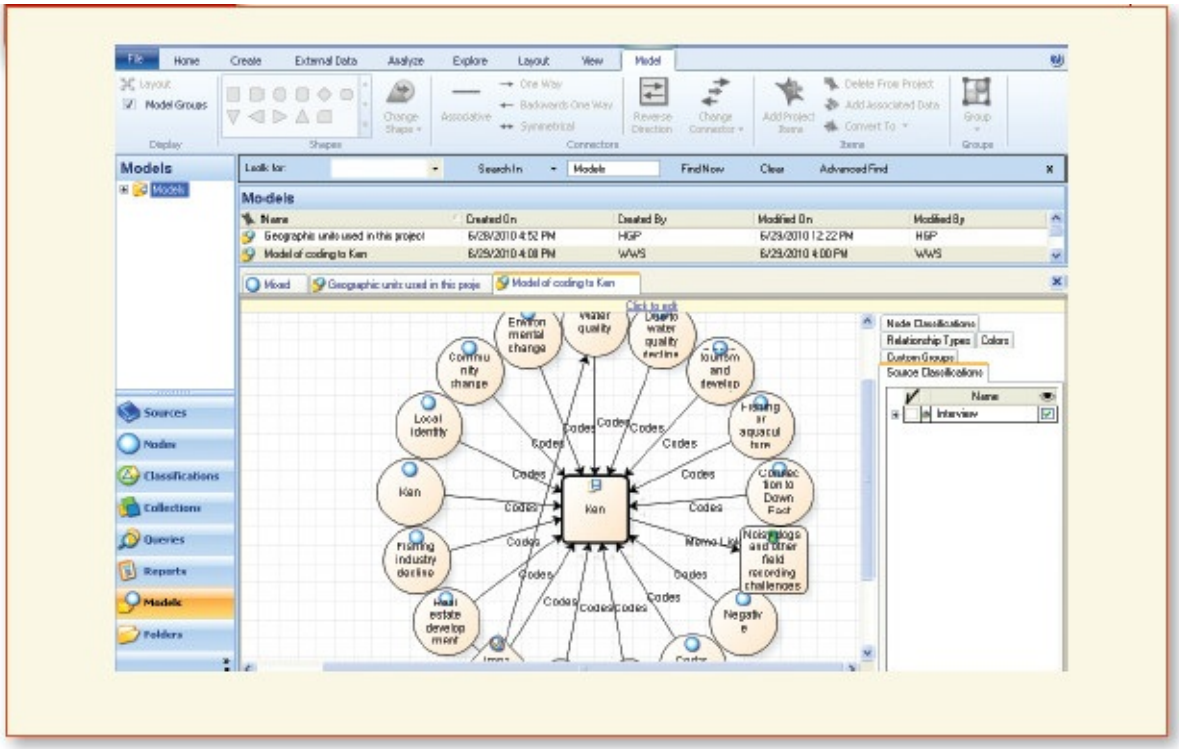

environmental change.nvp - NVivo
Nodes
Look for:
Search In
Nodes
Find Now
Clear
Advanced Find
Name
Sources
References
Created On
Created By
Modified On
Modified By
Attitude
Mixed
Negative
The economy has stopped the changes down here. There was a lot of proposed development in 2005, four years ago. There was a lot of – people were coming down here and trying to develop different areas, and the economy has stopped most of that. There was a group down here called Down East Tomorrow that was very vocal and followed a lot of the new subdivisions or communities or whatever, but, basically, the economy took care of that. Now that will probably come back if the building comes back and property values come back up. They'll probably have to fight that battle again, but we won't get into all that.
But this is a very fragile area. In 2005 I probably would have been on the other side. Being down here and being in the area and talking to commercial fishermen and meeting commercial fishermen and just people from the area – this area is so fragile that it can't stand a lot of –
Reference 2 - 6.39% Coverage
And the rules on that change all the time and they get harder and harder and more stringent. In the new rules, most every house you see down here – I say most – if there wasn't a septic tank on that lot there never would be one, because it won't, what they call 'perk.' It won't work. Not that it won't work. It's that in the new rules you gotta have an area for the house, for the well, and
Sources
Nodes
Classifications
Collections
Queries
Reports
Models

Exhibit 10.11 A Free-Form Model in NVivo



Decker and Van Winkle (1996) solved this problem by parenthetically inserting conceptual tags in the text whenever talk of drug sales was found. This process allowed them to examine all of the statements made by gang members about a single concept (drug sales). As you can imagine, however, this still left the researchers with many pages of transcript material to analyze.

Computer-assisted qualitative data analysis: Analysis of textual, aural, or pictorial data using a special computer program that facilitates searching and coding text.

What Ethical Issues Arise in Doing Qualitative Data Analysis?

The qualitative data analyst is never far from ethical issues and dilemmas. Throughout the analytic process, the analyst must consider how the findings will be used and how participants in the setting will react. Miles and Huberman (1994: 204–205) suggest several specific questions that should be kept in mind:

Journal Link

Read an article which uses triangulation to increase the study's validity.

Research integrity and quality.

Is my study being conducted carefully, thoughtfully, and correctly in terms of some reasonable set of standards? Real analyses have real consequences, so you owe it to yourself and those you study to adhere strictly to the analysis methods that you believe will produce authentic, valid conclusions.

Ownership of data and conclusions.

Who owns my field notes and analyses: I, my organization, my funders? And once my reports are written, who controls their dissemination? Of course, these concerns arise in any social research project, but the intimate involvement of the qualitative researcher with participants in the setting studied makes conflicts of interest between different stakeholders much more difficult to resolve. Working through the issues as they arise is essential.

Use and misuse of results.

Do I have an obligation to help my findings be used appropriately? What if they are used harmfully or wrongly? It is prudent to develop understandings early in the project with all major stakeholders that specify what actions will be taken to encourage the appropriate use of project results and to respond to what is considered misuse of these results.

Conclusion

The success of qualitative analyses may be difficult to judge, but Norman Denzin (2002) suggests that the following “interpretive criteria” questions could be asked:

- Does it illuminate the phenomenon as lived experience? In other words, do the materials bring the setting alive in terms of the people in that setting?
- Is it based on thickly contextualized materials? We should expect thick descriptions that encompass the social setting studied.
- Is it historically and relationally grounded? There must be a sense of the passage of time between events and the presence of relationships between social actors.
- Is the research processual and interactional? The researcher must have described the research process and his or her interactions within the setting.
- Does it engulf what is known about the phenomenon? This includes situating the analysis in the context of prior research and acknowledging the researcher’s own orientation upon first starting the investigation. (pp. 362–363)

If the answers are yes, a study has achieved much of the promise of qualitative research.

Key Terms

Computer-assisted qualitative data analysis 251
Emic focus 230
Etic focus 230
Grounded theory 244
Matrix 236
Narrative analysis 241
Progressive focusing 230
Qualitative data analysis 230
Tacit knowledge 239

Highlights

- Qualitative data analysis is guided by an emic focus of representing persons in the setting on their own terms, rather than by an etic focus on the researcher’s terms.
- Narrative analysis attempts to understand a life or a series of events as they unfolded in a meaningful progression.
- Conversation analysis studies the sequence and details of conversational interactions, primarily to understand how people construct social realities through their talk.
- Grounded theory connotes a general explanation that develops in interaction with the data and is continually tested and refined as data collection continues.
- Visual sociology uses the analysis of still photography and motion pictures (video, etc.) to learn both about society and about how people visualize their worlds.
- Special computer software can be used for the analysis of qualitative, textual, and pictorial data. Users can record their notes, categorize observations, specify links between categories, and count occurrences.
- Ethical issues in qualitative analysis often arise around how the results are used and how the subjects of the research may react to what has been done.

Student Study Site

The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. List the primary components of qualitative data analysis strategies. Compare and contrast each of these components with those relevant to quantitative data analysis. What are the similarities and differences? What differences do these make?
2. Does qualitative data analysis result in trustworthy results—in findings that achieve the goal of "authenticity"? Why would anyone question its use? What would you reply to the doubters?
3. Narrative analysis provides the "large picture" of how a life or event has unfolded, whereas conversation analysis focuses on the details of verbal interchange. When is each method most appropriate? How could one method add to the other?
4. Ethnography and grounded theory both refer to aspects of data analysis that are an inherent part of the qualitative approach. What do these approaches have in common? How do they differ? Can you identify elements of these two approaches in this chapter's examples of ethnomethodology, conversation analysis, and narrative analysis?

Finding Research

1. *The Qualitative Report* is an online journal about qualitative research. Inspect the table of contents for a recent issue (www.nova.edu/ssss/QR/index.html). Read one of the articles, and write a brief article review.
2. Be a qualitative explorer! Go to this list of qualitative research websites, and see what you can find that enriches your understanding of qualitative research (www.qualitativeresearch.uga.edu/QualPage/). Be careful to avoid textual data overload.

Critiquing Research

1. Read the complete text of one of the qualitative studies presented in this chapter, and evaluate its analysis and conclusions for authenticity, using the criteria in this chapter.

Doing Research

1. Attend a sports game as an ethnographer. Write up your analysis, and circulate it for criticism.
2. Write a narrative in class about your first date, car, college course, or something else you and your classmates agree on. Then collect all the narratives, and analyze them in a “committee of the whole.” Follow the general procedures discussed in the example of narrative analysis in this chapter.
3. Try out the HyperRESEARCH tutorials that you can link to on the book Study Site. How might qualitative analysis software facilitate the analysis process? Might it hinder the analysis process in some ways? Explain your answers.

Ethics Questions

1. Pictures are worth a thousand words, so to speak, but is that a thousand words too many? Should qualitative researchers (like yourself) feel free to take pictures of social interaction or other behaviors anytime, anywhere? What limits should an institutional review board place on researchers' ability to take pictures of others? What if the "after" picture of the Apache children in this chapter (Exhibit 10.7) also included Captain Pratt in a military uniform?
2. Participants in social settings often "forget" that an ethnographer is in their midst, planning to record what they say and do, even when the ethnographer has announced his role. New participants may not have heard the announcement, and everyone may simply get used to the ethnographer as if he were just "one of us." What efforts should an ethnographer take to keep people informed about his or her work in the setting under study? Consider settings such as a sports team, a political group, and a book group.

Video Interview Questions

1. Listen to the researcher interviews for Chapter 10 at edge.sagepub.com/chamblissmssw5e.
2. Paul Atkinson believes that researchers should think about not only what people are talking about but also "how" they are talking about a topic or concept. Do you agree with this statement? Why or why not?
3. What are his three suggestions for dealing with narratives?

Chapter 11 Unobtrusive Measures

©iStockphoto.com/peeterv

Learning Objectives

1. Define *unobtrusive measures*, and discuss their use in research, providing examples.
2. Describe the process of content analysis, and give one example.
3. Define both historical research methods and comparative research methods, and give an example of each.
4. Explain the process of event-structure analysis.
5. Identify the strengths and limitations of oral history.
6. Discuss the major methodological challenges that arise in comparative and historical research.
7. List some of the cautions and ethical issues to keep in mind when using unobtrusive methods.

Perhaps the most commonly used methods of social science research today are surveys (including political and opinion polling of all kinds) and face-to-face interviews. These methods can elicit tremendous amounts of valuable information, precisely tailored to the researcher's purposes, at a relatively low cost and with very little "dross," or irrelevant information. They can also use sophisticated sampling and create a close-up, human view on what is happening in social life.

But surveys and interviews have a great disadvantage: They are "**reactive**" methods in which the people being studied know they are being studied, and so may modify their answers or even the behavior being studied itself. Adult Americans routinely, for instance, overstate how much they vote, how much they exercise, and how frequently they attend church, whereas they underreport how frequently they tell lies. In an effect to offset the weaknesses of reactive measures, Eugene Webb and his colleagues (Webb et al. 1966; revised edition, 2000) assembled a wide variety of examples of what they called **unobtrusive measures**—that is, research techniques that would gather data without alerting the people under study. As Webb and company said, "So long as one has only a single class of data collection, and that class is the questionnaire or interview, one has inadequate knowledge." They urged that researchers use multiple methods in an effort to validate findings in various ways, and they put together a fascinating compendium of creative (some called them "oddball") ideas for studying social life: measuring interest in different museum exhibits by the

frequency with which floor tiles need to be replaced, discovering the most popular radio stations in town by having car mechanics note the settings on car radio dials, or glancing at the hands of patrons in a neighborhood bar to judge the level of manual work done by the patrons (calluses!).

Video Link

Watch a discussion on using unobtrusive measures in research.

Actually, there are many kinds of nonreactive research methods available. Webb at al. described four categories of data that might provide unobtrusive measures: physical traces, archives, simple observation, and contrived observation. We begin this chapter with a variety of examples of these more “creative” methods, mainly to suggest how broad these possibilities are. For the remainder of the chapter, we outline three far more commonly used important kinds of research that are also typically nonreactive: content analysis, historical research, and comparative analysis.

Unobtrusive measure: A measurement based on physical traces or other data that are collected without the knowledge or participation of the individuals or groups that generated the data.

Reactive methods: When the people being studied know they are being studied, and so may modify their answers or even the behavior being studied itself.

Creative Sources

Physical Traces

As criminal forensic scientists can attest, when human beings do almost anything, they tend to leave behind **physical traces** of themselves—hair, fingerprints, sweat, and such, certainly, but also wear and tear on the things they touch. Simply becoming aware of such traces (we might call it "seeing like a detective") can provide social scientists with valuable research data. On your way to class you might notice that the carpeting or tile on certain stairways is more worn than on others (as Webb suggested), that the chairs in some classrooms are more likely to be damaged, and that paper towels in one particular restroom always seems to run out first. These all these point to heavier traffic in some areas than in others, so that even without actually watching human beings moving you might be able estimate where they go. And your professor in class might well notice that some students' paperback books seem remarkably fresh, their backs uncracked and their pages unfilled with notes or underlinings; maybe students aren't doing the reading. Wear and tear on a book may only mean that it's a used book, but lack of such wear almost certainly suggests that no one, now or earlier, has read it.

Patterns of physical wear may change over time, revealing changes in usage. For instance, the famous tennis tournament at Wimbledon, in England, is played each year on grass courts, which of course will show usage more readily than would, say, a concrete court. Paul Kedrosky, an entrepreneur who thinks creatively about "data exhaust," or leftover sources of information, has suggested how in looking at photographs of a match from 25 years ago, you can see that the grass is worn in a pattern that moves up the middle of the court to the net (see Exhibit 11.1). The pattern shows how players would rush up to "volley" after their serves. But in the more recent photograph, the grass has been worn thin back at the baseline at the rear of the court, reflecting a "power baseline" game that has come to predominate in tennis currently.

Refuse, trash, even excretions of all sorts can be a fruitful source of information (as physicians have long known). "In December 2011, a pair of data collectors came to Boston. . . . They made 29 stops . . . walking the neighborhood streets and picking up discarded cigarette packs. They collected 253 packs in all," and by looking at the state excise tax stamp on each pack, determined that nearly 40% of cigarettes smoked in the Boston area were sold on the black market—

they had been illegally imported, to avoid the high cigarette taxes in the state (Hartnett 2014). And in one of the more creative uses of simple wastewater, “since all drug users urinate, and since the urine eventually winds up in the sewers, [Oregon State University chemist Jennifer] Field and her fellow researchers figured that sewer water would contain traces of whatever drugs the citizens were using.” Samples detected varying usage, by city, of cocaine, methamphetamine, and—most popular of all—caffeine. Cocaine use, interestingly, peaked on weekends, whereas methamphetamine use tended to hold steady across the week (Thompson 2007).

Exhibit 11.1 Patterns of Tennis Court Wear Showing Different Styles of Play

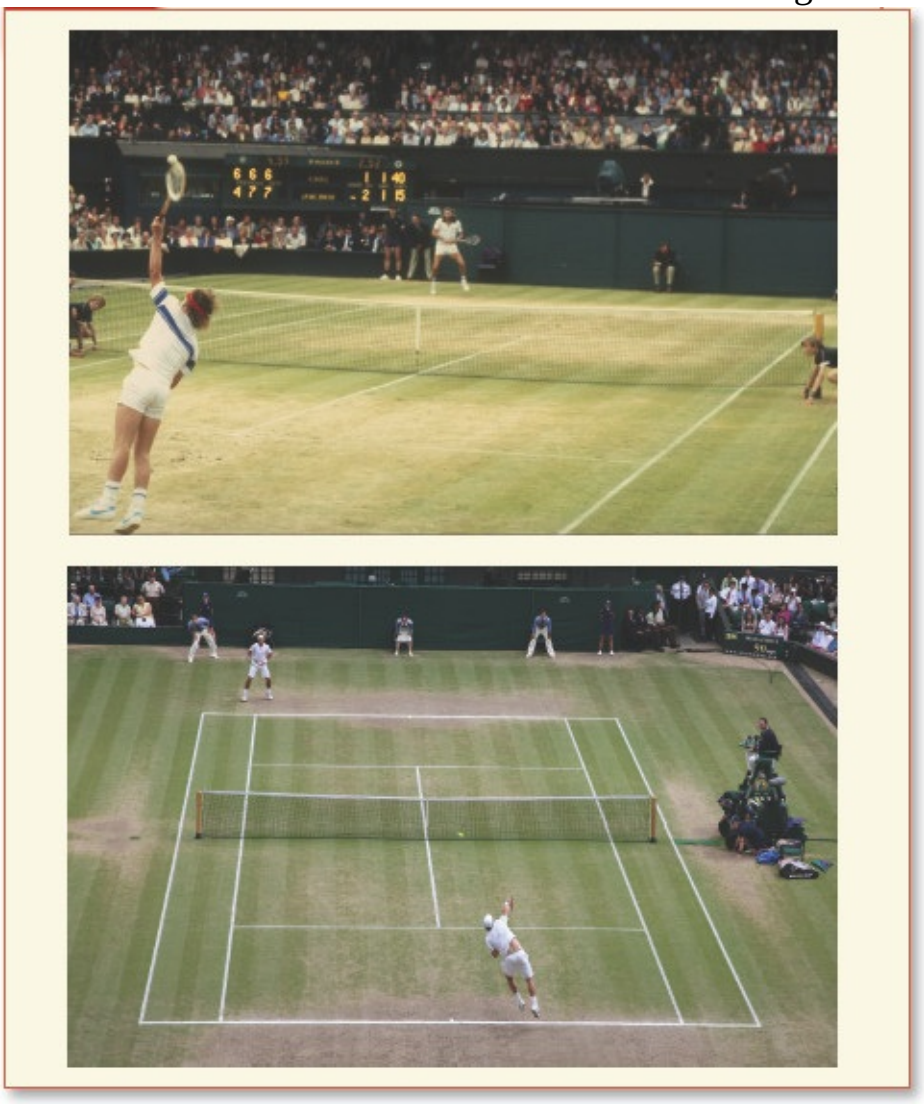


Photo Credits: ©Hugo Philpott/UPI/Newscom; Tony Duffy /Alls/Getty

Images.

Physical traces: Either the erosion or the accumulation of physical substances that can be used as evidence of activity. For instance, footprints in snow indicate that someone has walked there.

Archives

By *archives*, we just mean records of all sorts that are already being kept, aside from any social science purpose. These may be quite formal, as in government records of births, deaths, marriages, tax records, building permits, crime statistics, and the like. Law enforcement and health statistics provide, for example, a variety of community-level indicators of substance abuse (Gruenewald et al. 1997). Statistics on arrests for the sale and possession of drugs, drunk driving arrests, and liquor law violations (such as sales to minors) can usually be obtained on an annual basis, and often quarterly. Health-related indicators include single-vehicle fatal crashes, the rate of mortality from alcohol or drug abuse, and the use of treatment centers. All sorts of media create archives that can be mined for data, including newspapers, magazine articles, TV or radio talk shows, legal opinions, historical documents, personal cards and letters, diaries, or e-mail messages. Or one could learn about different U.S. cities, for instance, by looking at the "yellow book" business telephone directories that are still used by many establishments. You would discover there that Sarasota, Florida, has many pages devoted to nursing homes and hospital appliances, but Chattanooga, Tennessee, with roughly the same number of people, has fewer facilities for older people but a huge number of family-friendly churches.

Audio Link

Listen to a researcher describe the process of using archival research.

Less obvious, or even totally unintentional, archival sources can also be useful. An abandoned juvenile detention facility was the site, for John M. Klofas and Charles R. Cutshall (1985), of 2,765 instances of graffiti, in settings from the orientation corridors to inmates' cells to the bathrooms. The authors classified the graffiti by a number of variables including location and content, and concluded that juveniles upon entry seemed more concerned with establishing their individual identity and status, then later their concerns shifted to group affiliations. Romance, politics, and criticisms of the criminal justice system also figured prominently in what inmates wrote about (on the walls!). Archives of various sorts can also serve as a check on respondents' self-reports in surveys or

in interviews. In Michael Pollan's best-selling book *In Defense of Food* (2008), he first states that "validation studies' of dietary trials like the Women's Health Initiative or the Nurses's [Health] Study [conducted on more than 100,000 women over several decades] . . . indicate that people on average eat between a fifth and a third more than they say they do on questionnaires." He then adds, in a footnote, that "in fact, the magnitude of the error could be much greater, judging by the disparity between the total number of food calories produced every day for each American (3,900) and the average number of those calories Americans own up to chomping each day: 2,000. Waste can account for some of the disparity, but not nearly all of it" (Pollan 2008: 74).

With the proliferation of smartphones and handheld video recorders, photographic data has become far more available, providing archives of all sorts of routine as well as extraordinary historic events. The Japanese tsunami of 2011 was exceptionally well-documented, with real-time recordings of the wave as it came in, as water levels rose, and as the destruction ensued. As recently as the year 2000, almost no such evidence was easily available for study, but now even unpredicted tsunamis, tornadoes, flashfloods, and other catastrophes can and are being fully documented by people on the scene. YouTube and other video websites are a wonderful source for such recordings.

Photography itself has long provided valuable archival research material. Randall Collins, in research for his sweeping study *Violence: A Micro-Sociological Theory* (2008), assembled many hundreds of photos of people in violent situations from bank robberies to wartime combat to street riots. Collins's book is valuable methodologically for his detailed descriptions of how he selected photos, the sampling and interpretations involved, and the limitations of such data. Even given those issues, though, he was able to conclude (among many other important points) that in groups, violent activity tends to be confined to a few leaders—for instance in a riot in which a handful of protestors actually throw rocks while many more participants are just supportive or even passive (see Exhibit 11.2).

Archival data can be enormously useful, but as always you should be aware in using all sorts of archives that they may not accurately sample or represent reality. Even officially kept records, not to say personal documents, often have built-in biases. For instance, the level of blood alcohol legally required to establish intoxication can vary among communities, creating an appearance of different rates of abuse even though drinking and driving may in fact be similar.

Enforcement practices can vary as well among police jurisdictions, so that conclusions based on these records may be unjustified (Gruenewald et al. 1997: 14).

Exhibit 11.2 Leaders, Supporters, and Onlookers in a Riot

Photo Credit: MUSA AL-SHAER/AFP/Getty Images.

Archival data: Written or visual records, not produced by the researcher.

Observation

Of course, either moving or still photography is really just a recording of an observation—simply watching people. Fully developed, this is what we've called ethnography or field research (see Chapter 9), but even very brief observations can be revealing. Excellent work has been done, for instance, on the psychology of emotions, so that watching a person's face for even a fraction of a second can often tell you what they're feeling. Paul Ekman, a psychologist who has helped police forces establish when a suspect is lying or telling the truth (by their facial expressions), is an expert at making detailed observations of the facial features associated with different emotions. Here, in a tragic situation, Ekman describes the look on the face of a woman just told that her missing child has been found murdered (see Exhibit 11.3).

"One very strong and reliable sign [of intense sadness] is the angling upward of the inner corners of her eyebrows. It is reliable because few people can make this movement voluntarily, so it could rarely be deliberately fabricated. . . . Even when people are attempting not to show how they are feeling, these obliquely positioned eyebrows will often leak their sadness. Look at the space between her eyebrows. In most people a vertical wrinkle between the brows will appear, as it does here" (Ekman 2003: 97). A person well trained in Ekman's methods could do fascinating studies of different groups of people in public, following their emotional responses to various events. (Sporting events? Parties? Weddings?)

Research|Social Impact Link

Read about a study that uses observation to understand more about Autistic children.

Even simple and obvious sorts of observations, though, can be used to validate other sorts of measures. The tiny Scandinavian island nation of Iceland has very low official crime rates, according to standard police measures. But even casual observation suggests the same conclusion: It is common, for instance, to see babies in strollers lined up outside stores in Reykjavik, the capital, while mothers are inside shopping, a practice unthinkable (or even illegal, as parental negligence) in the United States. When Dan Chambliss lived in Iceland, at night

he saw children as young as 6 years old walking alone in downtown Reykjavik, and young women, obviously drunk, staggering home alone from dance clubs—what would be obviously quite dangerous in an American city was a perfectly safe, if perhaps embarrassing, practice in this benign environment.

Exhibit 11.3 Intense Emotion, Apparent by Close Observation

Source: Ekman, Paul. 2003. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. New York: Henry Holt.

At a far more complex level of “observation” stands the massive surveillance programs recently unveiled by the Edward Snowden leaks, in which the U.S. government was discovered to have been monitoring literally millions of telephone records, as well as hacking the intelligence services of other countries. Our computer-based lives are essentially being observed all the time, of course —by online providers, eager to see what we watch and click on, as well as by employers, who frequently keep track of e-mail and websurfing.

Contrived Observation

Sometimes researchers with access with online usage data carry out what Webb et al. called "**contrived observation**," that is, observation in which the researchers deliberately intervene in the observed activity—for instance, by experimenting. In June 2014, Facebook "revealed that it had manipulated the news feeds of over half a million randomly selected users to change the number of positive and negative posts they saw" (Goel 2014). Investigating the concern that perhaps seeing positive content posted by friends will make viewers feel negative or left out, the researchers (including academics as well as Facebook employees) deliberately modified what was shown on users' news feeds, to see how users would react. It turns out that people who see more positive content then produce more positive posts themselves. Facebook never asked explicit permission from the people who were studied (there were 689,003), although the company said that the 1.28 billion users give blanket permission when they begin using the service.

A more traditional form of contrived observation would be the groundbreaking linguistic field experiments conducted in the 1960s by William Labov, who hypothesized that people of different social classes pronounced their words differently (Labov 1972). (Specifically, Labov was curious about the way working-class residents of New York City sometimes drop their *r*s in casual conversation: "Hey, come over hee-ah!" instead of "over here!" might be an example.) If he used scheduled research interviews, Labov realized, subjects would speak more formally, but he wanted to find out how people pronounce their words in daily life, when they have no idea that they're being studied.

So Labov sent his research team members into three different New York City department stores (very popular in the 1960s), each representing a different social stratum of the city, as determined by various measures (prices, advertising budgets, etc.). Saks, on the upper East Side, was the expensive store, catering to an upper-class clientele; Macy's, at Herald Square, was somewhat more middle class; and S. Klein, now closed, was more a budget-level store. Assuming that sales people would to some extent mirror the accents of their customers, researchers would approach employees in each store and ask for directions to items they knew were stocked on the fourth floor of the building. Notice: "fourth floor," as a response, will provide two different uses of the letter *r*; when the

researcher would ask for clarification, the responding sales person would then emphasize the words clearly—giving in total, then, four different examples of the *r* sound. Labov and his team asked 264 subjects for the directions, and found that indeed, the more "upper crust" the store, the more likely the letter *r* was to be clearly sounded out—thus confirming his hypothesis of what Labov called "stylistic stratification." It was an excellent example of a contrived observation.

Contrived observation: Observations of situations in which the researcher has deliberately intervened.

Content Analysis

Certain forms of archival observation have been systematically developed into what's called **content analysis**. How are medical doctors regarded in U.S. culture? Do newspapers use the term *schizophrenia* in a way that reflects what this serious mental illness actually involves? Does the portrayal of men and women in video games reinforce gender stereotypes? Are the body images of male and female college students related to their experiences with romantic love? Content analysis, "the systematic, objective, quantitative analysis of message characteristics," is a method particularly well suited to the study of popular culture and many other issues concerning human communication (Neuendorf 2002: 1).

Content analysis develops inferences from human communication in any of its forms, including books, articles, magazines, songs, films, and speeches (Weber 1990: 9). This method was first applied to the study of newspaper and film content and then developed systematically for the analysis of Nazi propaganda broadcasts in World War II. Since then, content analysis has been used to study historical documents, records of speeches, and other "voices from the past" as well as media of all sorts (Neuendorf 2002: 31–37). The same techniques can now be used to analyze blog sites, wikis, and other text posted on the Internet (Gaiser & Schreiner 2009: 81–90).

The steps in a content analysis are represented in the flowchart in Exhibit 11.4. Note that the steps are comparable to the procedures in quantitative survey research. Use this flowchart as a checklist when you design or critique a content analysis project.

Kimberly Neuendorf's (2002: 3) analysis of medical prime time network television programming introduces the potential of content analysis. As Exhibit 11.5 shows, medical programming has been dominated by noncomedy shows, but there have been two significant periods of comedy medical shows—during the 1970s and early 1980s and then again in the early 1990s. It took a qualitative analysis of medical show content to reveal that the 1960s shows represented a very distinct "physician-as-God" era, which shifted to a more human view of the medical profession in the 1970s and 1980s. This era has been followed, in turn, by a mixed period that has had no dominant theme.

Identify a Population of Documents or Other Textual Sources

Perhaps the population will be all newspapers published in the United States, college student newspapers, nomination speeches at political party conventions, or “state of the nation” speeches by national leaders. Books or films are also common sources for content analysis projects. For her analysis of prime time programming since 1951, Neuendorf (2002: 3–4) used a published catalog of all TV shows. For Russ Schutt’s analysis with Duckworth and others (2003: 1402) of newspapers’ use of the terms *schizophrenia* and *cancer*, they requested a sample of articles from the LexisNexis national newspaper archive. Matthias Gerth and Gabriele Siegert (2012) focused on TV and newspaper stories during a 14-week Swiss political campaign, and Karen Dill and Kathryn Thill (2007: 855–856) turned to video game magazines for their analysis of the depiction of gender roles in video games. For their analysis of gender differences in body image and romantic love, Suman Ambwani and Jaine Strauss (2007: 15) surveyed students at a small midwestern liberal arts college.

Journal Link

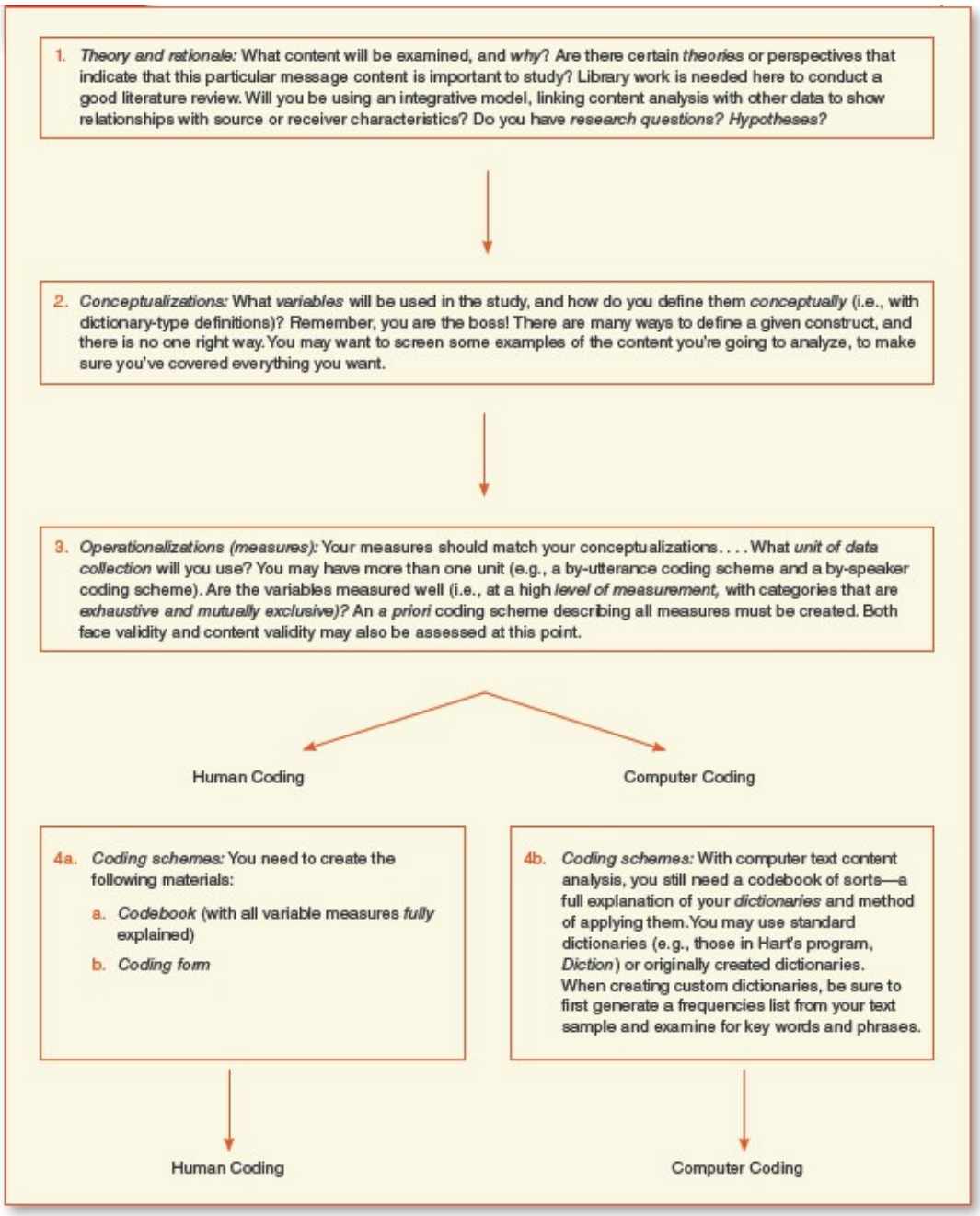
Read an article that uses content analysis to examine how race impacts social policy.

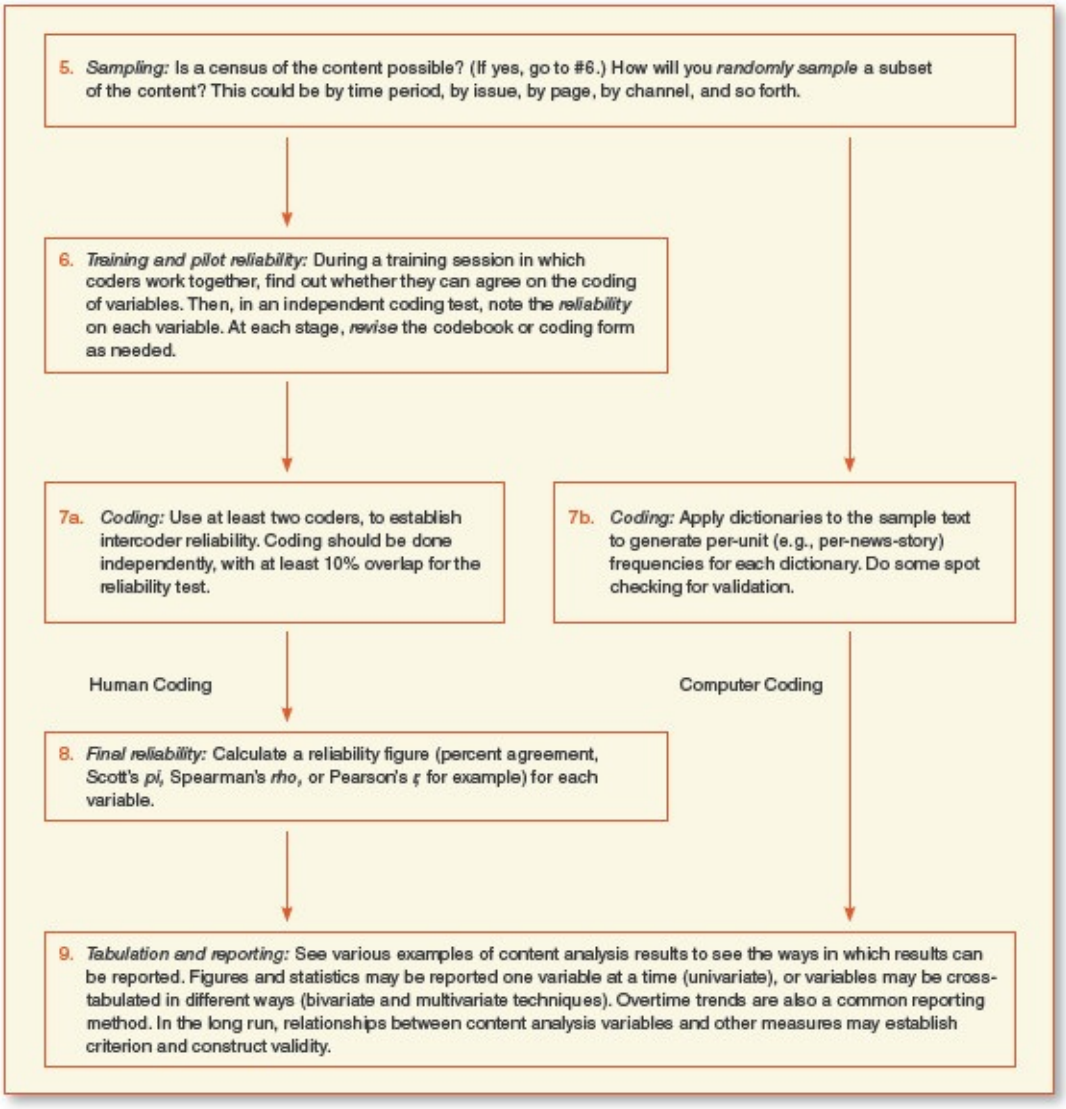
Content analysis: A research method for systematically analyzing and making inferences from text.

Determine the Units of Analysis

These could be items such as newspaper articles, whole newspapers, speeches, or political conventions, or they could be more microscopic units such as words, interactions, time periods, or other bits of a communication (Neuendorf 2002: 71). The units of analysis for Neuendorf (2002: 2) were “the individual medically oriented TV program”; for Duckworth et al. (2003: 1403), they were newspaper articles; for Gerth and Siegert (2012: 288) they were arguments made in media stories; and for Dill and Thill (2007: 856) they were images appearing in magazine articles. The units of analysis for Ambwani and Strauss (2007: 15) were individual students.

Exhibit 11.4 Flowchart for the Typical Process of Content Analysis Research


1. Theory and rationale: What content will be examined, and why? Are there certain theories or perspectives that indicate that this particular message content is important to study? Library work is needed here to conduct a good literature review. Will you be using an integrative model, linking content analysis with other data to show relationships with source or receiver characteristics? Do you have research questions? Hypotheses?
2. Conceptualizations: What variables will be used in the study, and how do you define them conceptually (i.e., with dictionary-type definitions)? Remember, you are the boss! There are many ways to define a given construct, and there is no one right way. You may want to screen some examples of the content you're going to analyze, to make sure you've covered everything you want.
3. Operationalizations (measures): Your measures should match your conceptualizations. . . . What unit of data collection will you use? You may have more than one unit (e.g., a by-utterance coding scheme and a by-speaker coding scheme). Are the variables measured well (i.e., at a high level of measurement, with categories that are exhaustive and mutually exclusive)? An a priori coding scheme describing all measures must be created. Both face validity and content validity may also be assessed at this point.
Human Coding
Computer Coding
4a. Coding schemes: You need to create the following materials:
a. Codebook (with all variable measures fully explained)
b. Coding form
4b. Coding schemes: With computer text content analysis, you still need a codebook of sorts—a full explanation of your dictionaries and method of applying them. You may use standard dictionaries (e.g., those in Hart's program, Diction) or originally created dictionaries. When creating custom dictionaries, be sure to first generate a frequencies list from your text sample and examine for key words and phrases.
Human Coding
Computer Coding

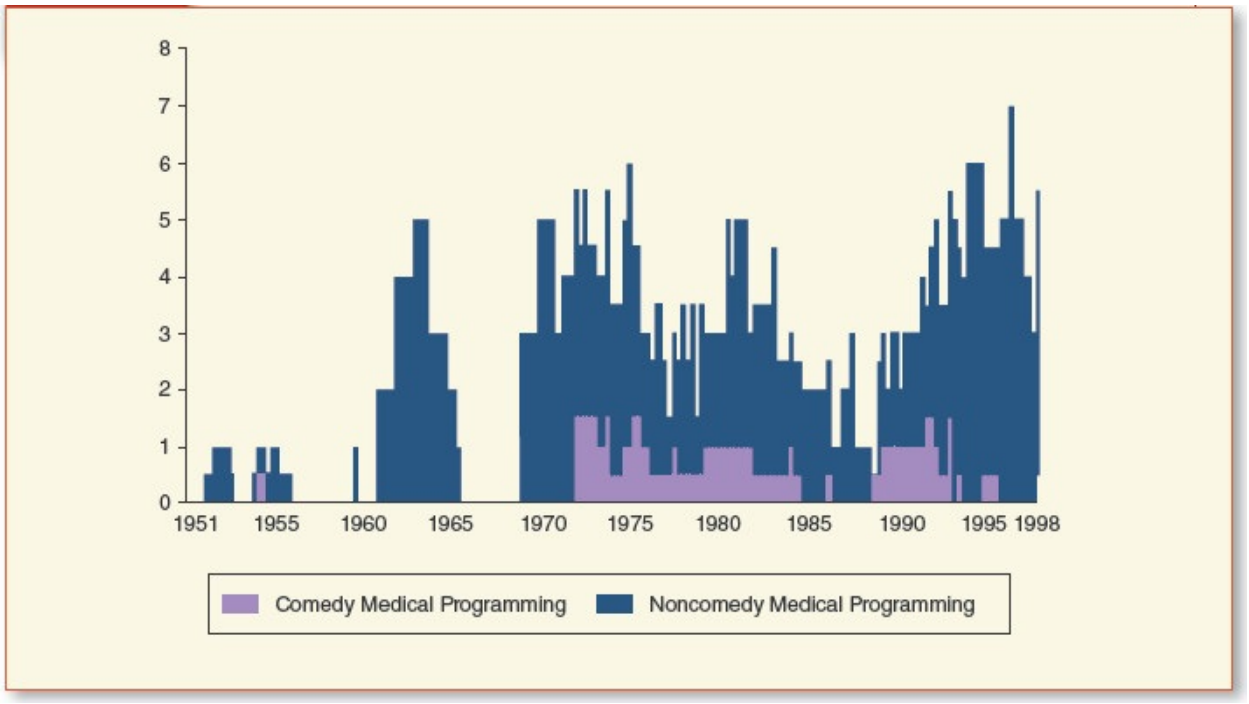

5. Sampling: Is a census of the content possible? (If yes, go to #6.) How will you randomly sample a subset of the content? This could be by time period, by issue, by page, by channel, and so forth.
6. Training and pilot reliability: During a training session in which coders work together, find out whether they can agree on the coding of variables. Then, in an independent coding test, note the reliability on each variable. At each stage, revise the codebook or coding form as needed.
7a. Coding: Use at least two coders, to establish intercoder reliability. Coding should be done independently, with at least 10% overlap for the reliability test.
7b. Coding: Apply dictionaries to the sample text to generate per-unit (e.g., per-news-story) frequencies for each dictionary. Do some spot checking for validation.
Human Coding
Computer Coding
8. Final reliability: Calculate a reliability figure (percent agreement, Scott's pi, Spearman's rho, or Pearson's r for example) for each variable.
9. Tabulation and reporting: See various examples of content analysis results to see the ways in which results can be reported. Figures and statistics may be reported one variable at a time (univariate), or variables may be cross-tabulated in different ways (bivariate and multivariate techniques). Overtime trends are also a common reporting method. In the long run, relationships between content analysis variables and other measures may establish criterion and construct validity.

Source: Neuendorf, Kimberly A. 2002. The content analysis guidebook. Thousand Oaks, CA: Sage.

Select a Sample of Units From the Population

The simplest strategy might be a simple random sample of documents. However, a stratified sample might be needed to ensure adequate representation of community newspapers in large and in small cities, or of weekday and Sunday papers, or of political speeches during election years and in off years (Weber 1990: 40–43). Nonrandom sampling methods have also been used in content analyses (Neuendorf 2002: 87–88).

Exhibit 11.5 Medical Prime Time Network Television Programming, 1951 to 1998



Content analysis typically proceeds according to a regular series of steps.

Source: Neuendorf, Kimberly A. 2002. The content analysis guidebook. Thousand Oaks, CA: Sage.

The selected samples in our exemplar content analysis projects were diverse. Neuendorf (2002: 2) included the entire population of medically oriented TV

programs between 1951 and 1998. For Schutt's content analysis with Duckworth (Duckworth et al. 2003), they had a student, Chris Gillespie, draw a stratified random sample of 1,802 articles published in the five U.S. newspapers with the highest daily circulation in 1996 to 1997 in each of the four regions identified in the LexisNexis database, as well as the two high-circulation national papers in the database, the *New York Times* and *USA Today* (pp. 1402–1403).

Because individual articles cannot be sampled directly in the LexisNexis database, a random sample of days was drawn first. All articles using the terms *schizophrenia* or *cancer* (or several variants of these terms) were then selected from the chosen newspapers on these days. Gerth and Siegert (2012: 285) selected 24 different newspapers and 5 TV news programs that targeted the population for the campaign, and then coded 3,570 arguments made in them about the campaign during its 14 weeks. Dill and Thill (2007: 855–856) used all images in the current issues (as of January 2006) of the six most popular video game magazines sold on Amazon.com. Ambwani and Strauss (2007: 15) used an availability sampling strategy, with 220 students from introductory psychology and a variety of other sources.

Design Coding Procedures for the Variables to Be Measured

This requires deciding what variables to measure, using the unit of text to be coded such as words, sentences, themes, or paragraphs. Then, the categories into which the text units are to be coded must be defined. These categories may be broad such as *supports democracy* or narrow such as *supports universal suffrage.* Development of clear instructions and careful training of coders is essential.

As an example, Exhibit 11.6 is a segment of the coding form that Schutt developed for a content analysis of union literature that he collected during a mixed-methods study of union political processes (Schutt 1986). His sample was of 362 documents: all union newspapers and a stratified sample of union leaflets given to members during the years of the investigation. The coding scheme included measures of the source and target for the communication, as well as measures of concepts that the theoretical framework indicated were important.

The analysis showed a decline in concern with client issues and an increase in focus on organizational structure.

Developing reliable and valid coding procedures is not an easy task. The meaning of words and phrases is often ambiguous. Coding procedures cannot simply categorize and count words; text segments in which the words are embedded must also be inspected before codes are finalized. Because different coders may perceive different meanings in the same text segments, explicit coding rules are required (Weber 1990: 23–29).

After coding procedures are developed, their reliability should be assessed by comparing different coders' codes for the same variables. Computer programs for content analysis can enhance reliability by facilitating the consistent application of text-coding rules (Weber 1990: 24–28). Validity can be assessed with a construct validation approach by determining the extent to which theoretically predicted relationships occur (see Chapter 4).

Neuendorf's (2002: 2) analysis of medical programming measured two variables that did not need explicit coding rules: length of show in minutes and the year(s)

the program was aired. She also coded shows as comedies or noncomedies, as well as medical or not.

Research|Social Impact Link

Read about how content was coded in looking at the history of women at the *New York Times.*

Dill and Thill (2007) used two coders and a careful training procedure for their analysis of the magazine images about video games:

> One male and one female rater, both undergraduate psychology majors, practiced on images from magazines similar to those used in the current investigation. Raters discussed these practice ratings with each other and with the first author until they showed evidence of properly applying the coding scheme for all variables. Progress was also checked part way through the coding process, as suggested by [Gloria] Cowan (2002). Cowan (2002) reports that this practice of reevaluating ratings criteria is of particular value when coding large amounts of violent and sexual material because, as with viewers, coders suffer from desensitization effects. (Dill & Thill 2007: 856)

Develop Appropriate Statistical Analyses

The content analyst creates variables for analysis by counting occurrences of particular words, themes, or phrases and then tests relations between the resulting variables. These analyses could use some of the statistics that were introduced in Chapter 8, including frequency distributions, measures of central tendency and variation, cross-tabulations, and correlation analysis (Weber 1990: 58–63). Computer-aided qualitative analysis programs can help, in many cases, develop coding procedures and then to carry out the content coding.

The simple chart that Neuendorf (2002: 3) used to analyze the frequency of medical programming appears in Exhibit 11.5. Schutt's content analysis with Duckworth and others (2003) was simply a comparison of percentages showing that 28% of the articles mentioning schizophrenia used it as a metaphor, compared with only 1% of the articles mentioning cancer. We also presented examples of the text that had been coded into different categories. For example, *the nation's schizophrenic perspective on drugs* was the type of phrase coded as a metaphorical use of the term *schizophrenia* (p. 1403). Dill and Thill (2007: 858) presented percentages and other statistics that showed that, among other differences, female characters were much more likely to be portrayed in sexualized ways in video game images than were male characters. Ambwani and Strauss (2007: 16) used other statistics that showed that body esteem and romantic love experiences are related, particularly for women. They also examined the original written comments and found further evidence for this relationship. For example, one woman wrote, "[My current boyfriend] taught me to love my body. Now I see myself through his eyes, and I feel beautiful" (p. 17). Content analysis, then, has the power to reveal broad patterns in how people understand even the most intimate sorts of experiences.

Exhibit 11.6 Union Literature Coding Form*

I. Preliminary Codes

1. Document # _____
2. Date _____

   mo yr
3. Length of text _____ pp. (round up to next 1/4 page; count legal size as 1.25)
4. Literature Type
   1. General leaflet for members/employees
   2. Newspaper/Newsletter article
   3. Rep Council motions
   4. Other material for Reps, Stewards, Delegates (e.g., budget, agenda)
   5. Activity reports of officers, President's Report
   6. Technical information-filing grievances, processing forms
   7. Buying plans/Travel packages
   8. Survey Forms, Limited Circulation material (correspondence)
   9. Non-Union
   10. Other _____ (specify)

4A. If newspaper article

Position

1. Headline story
2. Other front page
3. Editorial
4. Other

4B. If Rep Council motion

Sponsor

1. Union leadership
2. Office
3. Leadership faction
4. Opposition faction
5. Other

5. Literature content—Special issues
   1. First strike (1966)
   2. Second strike (1967)
   3. Collective bargaining (1977)
   4. Collective bargaining (1979)
   5. Election/campaign literature
   6. Affiliation with AFSCME/SEIU/other national union
   7. Other

II. Source and Target

6. Primary source (code in terms of those who prepared this literature for distribution).
   1. Union-newspaper (Common Sense; IUPAE News)
   2. Union-newsletter (Info and IUPAE Bulletin)
   3. Union-unsigned
   4. Union officers
   5. Union committee
   6. Union faction (the Caucus; Rank-and-Filers; Contract Action, other election slate; PLP News; Black Facts)
   7. Union members in a specific work location/office
   8. Union members—other
   9. Dept. of Public Aid/Personnel
   10. DVR/DORS

11. Credit Union
12. Am. Buyers' Assoc.
13. Other nonunion

7. Secondary source (use for lit. at least in part reprinted from another source, for distribution to members)
   1. Newspaper—general circulation
   2. Literature of other unions, organizations
   3. Correspondence of union leaders
   4. Correspondence from DPA/DVR-DORS/Personnel
   5. Correspondence from national union
   6. Press release
   7. Credit Union, Am. Buyers'
   8. Other _____ (specify)
   9. None

8. Primary target (the audience for which the literature is distributed)
   1. Employees—general (if mass-produced and unless otherwise stated)
   2. Employees-DVR/DORS
   3. Union members (if refers only to members or if about union elections)
   4. Union stewards, reps, delegates committee
   5. Non-unionized employees (recruitment lit, etc.)
   6. Other _____ (specify)
   7. Unclear

III. Issues

A. Goal
B. Employee conditions/benefits (Circle up to 5)
   1. Criteria for hiring
   2. Promotion
   3. Work out of Classification, Upgrading
   4. Step increases
   5. Cost-of-living, pay raise, overtime pay, "money"
   6. Layoffs (nondisciplinary); position cuts
   7. Workloads, Redeterminations, "30 for 40," GA Review
   8. Office physical conditions, safety
   9. Performance evaluations
   10. Length of workday
   11. Sick benefits/leave—holidays, insurance, illness, vacation, voting time
   12. Educational leave
   13. Grievances—change in procedures
   14. Discrimination (race, sex, age, religion, national origin)
   15. Discipline—political (union-related)
   16. Discipline—performance, other
   17. Procedures with clients, at work
   18. Quality of work, "worthwhile jobs"—other than relations with clients

*Coding instructions available from author.

Source: Reprinted by permission from Schutt, Russell K. 1986. Organization in a changing environment. Albany: State University of New York Press. Reprinted by permission of The State University of New York Press. All rights reserved.

Historical Methods

The central insight behind both historical and comparative research, as we will see, is that we can improve our understanding of social process when we make comparisons with other times and places. Max Weber's comparative study of world religions (Bendix 1962) and Émile Durkheim's (1984) historical analysis of the division of labor are two examples of the central role of historical and comparative research during the period sociology emerged as a discipline. Although the popularity of this style of research ebbed with the growth of survey methods and statistical analysis in the 1930s, exemplary works such as Reinhard Bendix's (1956) *Work and Authority in Industry* and Barrington Moore Jr.'s (1966) *Social Origins of Democracy and Dictatorship* helped fuel a resurgence of historical and comparative methods in the 1970s and 1980s that has continued into the 21st century (Lange 2013: 22–33). In recent years, the globalization of U.S. economic ties and the internationalization of scholarship have increased the use of unobtrusive methods for comparative research across many different countries (Kotkin 2002).

Historical methods are used increasingly by social scientists in sociology, anthropology, political science, and economics, as well as by many historians (Monkkonen 1994). The late 20th and early 21st centuries have seen so much change in so many countries that many scholars have felt a need to investigate the background of these changes and to refine their methods of investigation (Hallinan 1997; Robertson 1993).

Much historical research is qualitative. Like other qualitative methods, qualitative historical research is inductive: it develops an explanation for what happened from the details discovered about the past. In addition, qualitative historical research is **case-oriented**; it focuses on the nation or other unit as a whole, rather than only on different parts of the whole in isolation from each other (Ragin 2000: 68). The research question is "What was Britain like at the time?" rather than "What did Queen Elizabeth do?" Related to this case orientation, qualitative historical research is **holistic**—concerned with the context in which events occurred and the interrelations between different events and processes: "how different conditions or parts fit together" (Ragin 1987: 25–26). Charles Ragin (2000: 67–68) uses the example of case-oriented research on the changing relationship between income and single parenthood in the United

States after World War II:

> In the end, the study is also about the United States in the second half of the twentieth century, not just the many individuals and families included in the analysis. More than likely, the explanation of the changing relation between income and single parenthood would focus on interrelated aspects of the United States over this period. For example, to explain the weakening link between low income and single parenthood the researcher might cite the changing status of women, the decline in the social significance of conventional family forms, the increase in divorce, the decrease in men's job security, and other changes occurring in the United States over this period.

Qualitative historical research is also likely to be *historically specific*—limited to the specific time(s) and place(s) studied. Qualitative historical research uses **narrative explanations**—in which the research tells a story involving specific actors and other events occurring at the same time (Abbott 1994: 102) or one that accounts for the position of actors and events in time and in a unique historical context (Griffin 1992). Larry Griffin's (1993) research on lynching, in the next section, provides a good example.

The focus on the past presents special methodological challenges:

- Documents and other evidence may have been lost or damaged.
- Available evidence may represent a sample biased toward more newsworthy figures.
- Written records will be biased toward those who were more prone to writing.
- Feelings of individuals involved in past events may be hard, if not impossible, to reconstruct.

Before you judge historical social science research as credible, you should look for convincing evidence that each of these challenges has been addressed.

Event-Structure Analysis

One technique useful in historical research is **event-structure analysis**. Event-structure is a qualitative approach that relies on a systematic coding of key events or national characteristics to identify the underlying structure of action in a chronology of events. The codes are then used to construct event sequences, make comparisons between cases, and develop an idiographic causal explanation for a key event.

An event-structure analysis consists of the following steps:

1. Classifying historical information into discrete events
2. Ordering events into a temporal sequence
3. Identifying prior steps that are prerequisites for subsequent events
4. Representing connections between events in a diagram
5. Eliminating from the diagram connections that are not necessary to explain the focal event

Griffin (1993) used event-structure analysis to explain a unique historical event, a lynching in the 1930s in Mississippi. According to published accounts and legal records, the lynching occurred after David Harris, an African American who sold moonshine from his home, was accused of killing a white tenant farmer. After the killing was reported, the local deputy was called and a citizen search party was formed. The deputy did not intervene as the search party trailed Harris and then captured and killed him. Meanwhile, Harris's friends killed another African American who had revealed Harris's hiding place. This series of events is outlined in Exhibit 11.7.

Which among the numerous events occurring between the time that the tenant farmer confronted Harris and the time that the mob killed Harris had a causal influence on that outcome? To identify these idiographic causal links, Griffin identified plausible counterfactual possibilities—events that might have occurred but did not—and considered whether the outcome might have been changed if a counterfactual had occurred instead of a particular event.

> If, contrary to what actually happened, the deputy had attempted to stop the mob, might the lynching have been averted? . . . Given what happened in

> comparable cases and the Bolivar County deputy's clear knowledge of the existence of the mob and of its early activities, his forceful intervention to prevent the lynching thus appears an objective possibility. (Griffin 1993: 1112)

So, Griffin concluded that nonintervention by the deputy had a causal influence on the lynching.

Audio Link

Listen to a project detailing the oral history of women in technology.

Case-oriented research: Research that focuses attention on the nation or other unit as a whole.

Holistic research: Research concerned with the context in which events occurred and the interrelations between different events and processes.

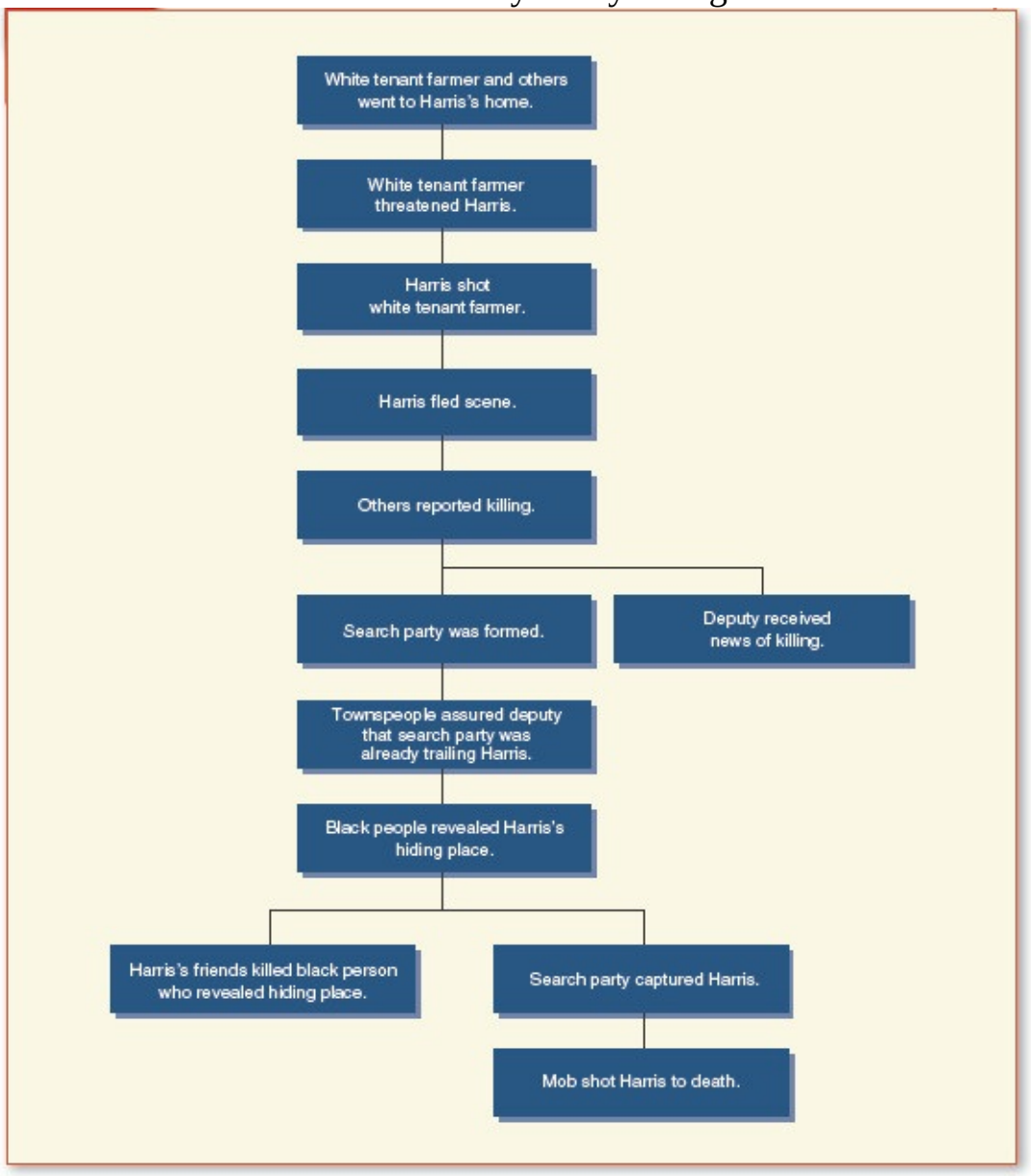
Narrative explanation: An explanation that involves developing a narrative of events and processes that indicate a chain of causes and effects.

Event-structure analysis: A systematic method of developing a causal diagram showing the structure of action underlying some chronology of events; the result is an idiographic causal explanation.

Oral History

History that is not written down is mostly lost to posterity (and social researchers). However,**oral histories** can be useful for understanding historical events that occurred within the lifetimes of living individuals. As the next example shows, sometimes oral histories even result in a written record that can be analyzed by researchers at a later point in time.

Exhibit 11.7 Event-Structure Analysis: Lynching Incident in the 1930s


White tenant farmer and others went to Harris's home.
White tenant farmer threatened Harris.
Harris shot white tenant farmer.
Harris fled scene.
Others reported killing.
Search party was formed.
Deputy received news of killing.
Townspeople assured deputy that search party was already trailing Harris.
Black people revealed Harris's hiding place.
Harris's friends killed black person who revealed hiding place.
Search party captured Harris.
Mob shot Harris to death.

Source: Griffin, Larry J. 1993. Narrative, event-structure analysis, and causal interpretation in historical sociology. American Journal of Sociology 98(March): 1110. Reprinted with permission from the University of Chicago Press.

Thanks to a Depression-era writers' project, Deanna Pagnini and Philip Morgan (1996) found that they could use oral histories to study attitudes toward births out of wedlock among African American and white women in the South during the 1930s.

Almost 70% of African American babies are born to unmarried mothers, compared with 22% of white babies (Pagnini & Morgan 1996: 1696). This difference often is attributed to contemporary welfare policies or problems in the inner city, but Pagnini and Morgan thought it might be the result of more enduring racial differences in marriage and childbearing. To investigate these historical differences, they read 1,170 life histories recorded by almost 200 writers who worked for a New Deal program during the Depression of the 1930s, the Federal Writers' Project Life History Program for the Southeast. The interviewers had used a topic outline that included family issues, education, income, occupation, religion, medical needs, and diet.

In 1936, the divergence in rates of nonmarital births was substantial in North Carolina: 2.6% of white births were to unmarried women, compared with 28.3% of nonwhite births. The oral histories gave some qualitative insight into community norms that were associated with these patterns. A white seamstress who became pregnant at age 16 recalled, "I'm afraid he didn't want much to marry me, but my mother's threats brought him around" (Pagnini & Morgan 1996: 1705). There were some reports of suicides by unwed young white women who were pregnant. In comparison, African American women who became pregnant before they were married reported regrets, but rarely shame or disgrace. There were no instances of young black women committing suicide or getting abortions in these circumstances.

> We found that bearing a child outside a marital relationship was clearly not the stigmatizing event for African Americans that it was for whites. . . . When we examine contemporary family patterns, it is important to remember that neither current marriage nor current childbearing patterns

> are “new” for either race. Our explanations for why African Americans and whites organize their families in different manners must take into account past behaviors and values. (Pagnini & Morgan 1996: 1714–1715)

Whether oral histories are collected by the researcher or obtained from an earlier project, the stories they tell can be no more reliable than the memories that are recalled. Unfortunately, memories of past attitudes are “notoriously subject to modifications over time” (Banks 1972: 67), as are memories about past events, relationships, and actions. Use of corroborating data from documents or other sources should be used when possible to increase the credibility of descriptions based on oral histories.

One common measurement problem in historical research projects is the lack of data from some historical periods (Rueschemeyer, Stephens, & Stephens 1992: 4; Walters, James, & McCammon 1997). For example, the widely used U.S. Uniform Crime Reporting System did not begin until 1930 (Rosen 1995). Sometimes, alternative sources of documents or estimates for missing quantitative data can fill in gaps (Zaret 1996), but even when measures can be created for key concepts, multiple measures of the same concepts are likely to be out of the question; as a result, tests of reliability and validity may not be feasible (Bollen, Entwisle, & Alderson 1993; Paxton 2002).

The available measures are not always adequate. What is included in the historical archives may be an unrepresentative selection of materials that remain from the past. At various times, some documents could have been discarded, lost, or transferred elsewhere for a variety of reasons. *Original* documents may be transcriptions of spoken words or handwritten pages and could have been modified slightly in the process; they could also be outright distortions (Erikson 1966: 172, 209–210; Zaret 1996). When relevant data are obtained from previous publications, it is easy to overlook problems of data quality, but this simply makes it all the more important to evaluate the primary sources.

Oral history: Data collected through intensive interviews with participants in past events.

Comparative Methods

The limitations of single-case historical research have encouraged many social scientists to turn to comparisons between nations. These studies allow for a broader vision about social relations than is possible with cross-sectional research limited to one country or other unit.

Cross-Sectional Comparative Research

Comparisons between countries during one time period can help social scientists identify the limitations of explanations based on single-nation research. Such comparisons can suggest the relative importance of universal factors in explaining social phenomena compared with unique factors rooted in specific times and places (de Vaus 2008: 251). These comparative studies may focus on a period in either the past or the present. Peter Houtzager and Arnab Acharya (2011) also point out that it can be more appropriate to compare cities or regions when the nations in which they are embedded vary internally in their social characteristics. For example, they compare the impact of engagement in associations on citizenship activity in São Paolo, Brazil, and Mexico City because the conditions exist for such an impact in these cities, rather than in the surrounding countries.

Historical and comparative research that is quantitative may obtain data from national statistics or other sources of published data; if it is contemporary, such research may rely on cross-national surveys. Like other types of quantitative research, quantitative historical and comparative research can be termed **variable-oriented research**, with a focus on variables representing particular aspects of the units studied (Demos 1998).

Causal reasoning in quantitative comparative research is nomothetic, and the approach is usually deductive, testing explicit hypotheses about relations between these variables (Kiser & Hechter 1991). For example, Clem Brooks and Jeff Manza (2006: 476–479) deduce from three theories about welfare states—national values, power resources, and path dependency theory—the hypothesis that voters' social policy preferences will influence welfare state expenditures. Using country-level survey data collected by the International Social Survey Program (ISSP) in 15 democracies in five different years and expenditure data from the Organisation for Economic Co-operation and Development (OECD), Brooks and Manza were able to identify a consistent relationship between popular preferences for social welfare spending and the actual national expenditures (see Exhibit 11.8).

Exhibit 11.8 Interrelationship of Policy Preferences and Welfare State Output

Note: Scattergram shows data for policy preferences and welfare state spending in 15 OECD democracies. Data are from the ISSP/OECD (International Social Survey Program/Organisation for Economic Co-operation and Development).

Source: Brooks, Clem, and Jeff Manza. June 2006. Social policy responsiveness in developed democracies. American Sociological Review 71(3): 474–494. Reprinted with permission from the American Sociological Association.

Journal Link

Read Brooks & Manza's study that utilizes variable-oriented research

Popular preferences are important factors in political debates over immigration policy. Christopher Bail (2008) asked whether majority groups in different European countries differ in the way that they construct "symbolic boundaries" that define "us" versus an immigrant "them." For his cross-sectional comparative investigation, he drew on 333,258 respondents in the 21-country European Social Survey (ESS). The key question about immigrants in the ESS was: "Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside [country] should be able to come and live here." The "things" whose importance they were asked to rate were six individual characteristics: being (1) white, (2) well educated, (3) from a Christian background, (4) speaking the official national language, (5) being committed to the country's way of life, and (6) having work skills needed in the country. Bail then calculated the average importance rating in each country for each of these characteristics and used a statistical procedure to cluster the countries by the extent to which their ratings and other characteristics were similar.

Bail's (2008: 54–56) analysis identified the countries as falling into three clusters (see Exhibit 11.9). Cluster A countries are on the periphery of Europe and have only recently experienced considerable immigration; their populations tend to draw boundaries by race and religion. Cluster B countries are in the core of Western Europe (except Slovenia), have a sizable and long-standing immigrant population, and their populations tend to base their orientations toward immigrants on linguistic and cultural differences. Countries in Cluster C are in Scandinavia, have a varied but relatively large immigrant population, and attach much less importance to any of the six symbolic boundaries than do those in the other countries. Bail (2008: 56) encourages longitudinal research to determine the extent to which these different symbolic boundaries are the product or the source of social inequality in these countries.

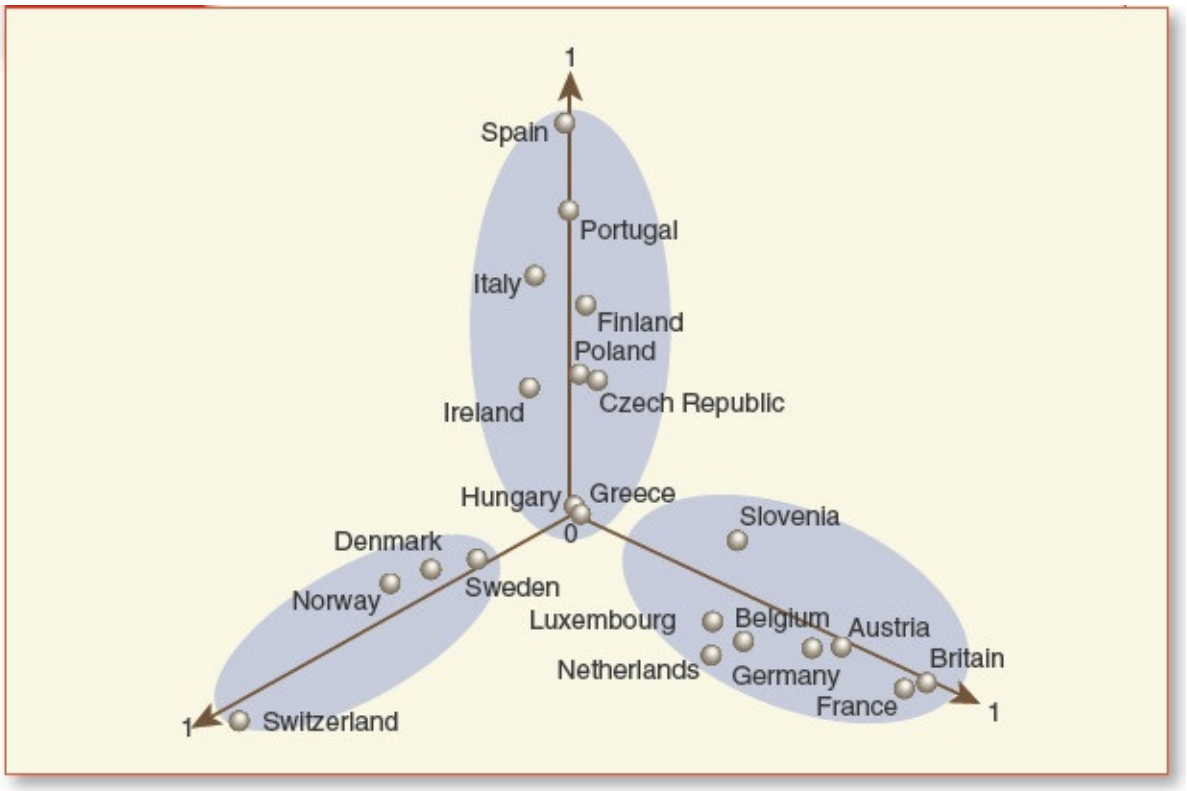
Cross-sectional comparative research has also helped explain variation in voter turnout. This research focuses on a critical issue in political science: Although free and competitive elections are a defining feature of democratic politics,

elections themselves cannot orient governments to popular sentiment if citizens do not vote (LeDuc, Niemi, & Norris 1996). As a result, the low levels of voter participation in U.S. elections have long been a source of practical concern and research interest.

International data give our first clue for explaining voter turnout: The historic rate of voter participation in the United States (48.3%, on average) is much lower than it is in many other countries that have free, competitive elections; for example, Italy has a voter turnout of 92.5%, on average, since 1945 (Exhibit 11.10).

Does this variation result from differences between voters in knowledge and wealth? Do media and political party get-out-the-vote efforts matter? Mark Franklin's (1996: 219–222) analysis of international voting data indicates that neither explanation accounts for much of the international variation in voter turnout. Instead, the structure of competition and the importance of issues are influential. Voter turnout is maximized where structural features maximize competition: compulsory voting (including, in Exhibit 11.10, Austria, Belgium, Australia, and Greece), mail and Sunday voting (including the Netherlands and Germany), and multiday voting. Voter turnout also tends to be higher where the issues being voted on are important and where results are decided by proportional representation (as in Italy and Israel, in Exhibit 11.10) rather than on a winner-take-all basis (as in U.S. presidential elections)—so individual votes are more important.

Exhibit 11.9 Symbolic Boundaries Against Immigrants in 21 European Countries



Source: Bail, Christopher A. February 2008. The configuration of symbolic boundaries against immigrants in Europe. American Sociological Review 73(1): 37–59. Reprinted with permission from the American Sociological Association.

Exhibit 11.10 Average Percentage of Voters Who Participated in Presidential or Parliamentary Elections, 1945–1998*

Country	Vote %	Country	Vote %
Italy	92.5	St. Kitts and Nevis	58.1
Cambodia	90.5	Morocco	57.6
Seychelles	96.1	Cameroon	56.3
Iceland	89.5	Paraguay	56.0
Indonesia	88.3	Bangladesh	56.0
New Zealand	86.2	Estonia	56.0
Uzbekistan	86.2	Gambia	55.8
Albania	85.3	Honduras	55.3
Austria	85.1	Russia	55.0
Belgium	84.9	Panama	53.4
Czech	84.8	Poland	52.3
Netherlands	84.8	Uganda	50.6
Australia	84.4	Antigua and Barbuda	50.2
Denmark	83.6	Burma/Myanmar	50.0
Sweden	83.5	Switzerland	49.3
Mauritius	82.8	USA	48.3
Portugal	82.4	Mexico	48.1
Mongolia	82.3	Peru	48.0
Tuvalu	81.9	Brazil	47.9
Western Samoa	81.9	Nigeria	47.6
Andorra	81.3	Thailand	47.4
Germany	80.9	Sierra Leone	46.8
Slovenia	80.6	Botswana	46.5
Aruba	80.4	Chile	45.9
Namibia	80.4	Senegal	45.6
Greece	80.3	Ecuador	44.7
Guyana	80.3	El Salvador	44.3
Israel	80.0	Haiti	42.9
Kuwait	79.6	Ghana	42.4
Norway	79.5	Pakistan	41.8
San Marino	79.1	Zambia	40.5
Finland	79.0	Burkina Faso	38.3
Suriname	77.7	Nauru	37.3
Malta	77.6	Yemen	36.8
Bulgaria	77.5	Colombia	36.2
Romania	77.2	Niger	35.6

*Based on entire voting-age population in countries that held at least two elections during these years. Only countries with highest and lowest averages are shown.

Source: Reproduced by permission of International IDEA from Turnout in the world—country by country performance (1945-1998). From Voter Turnout: A Global Survey (http://www. int/vt/survey/voter_turnout_pop2-2.cfm) © International Institute for Democracy and Electoral Assistance.

Franklin concludes that these characteristics explain the low level of voter turnout in the United States, rather than the characteristics of individual voters. The United States lacks the structural features that make voting easier, the

proportional representation that increases the impact of individuals' votes, and, often, the sharp differences between candidates that are found in countries with higher turnout. Because these structural factors generally do not vary within nations, we would never realize their importance if our analysis was limited to data from individuals in one nation.

Despite the unique value of comparative analyses like Franklin's (1996), such cross-national research also confronts unique challenges (de Vaus 2008: 255). The meaning of concepts and the operational definitions of variables may differ between nations or regions (Erikson 1966: xi), so the comparative researcher must consider how best to establish measurement equivalence (Markoff 2005: 402). For example, the concept of being a *good son or daughter* refers to a much broader range of behaviors in China than in most Western countries (Ho 1996). Rates of physical disability cannot be compared between nations because standard definitions are lacking (Martin & Kinsella 1995: 364–365). Individuals in different cultures may respond differently to the same questions (Martin & Kinsella 1995: 385). Alternatively, different measures may have been used for the same concepts in different nations, and the equivalence of these measures may be unknown (van de Vijver & Leung 1997: 9). The value of statistics for particular geographic units such as counties in the United States may vary over time simply because of changes in the boundaries of these units (Walters et al. 1997). Such possibilities should be considered, and any available opportunity should be taken to test for their effects.

Careers and Research

Kurt Taylor Gaubatz, PhD, Faculty,
Department of Political Science & Geography
International Studies Graduate Program

Source: Kurt Taylor Gaubatz

Kurt Taylor Gaubatz is the quintessential comparative researcher whose book *Elections and War*

(1999) exemplifies the approach. But he started college at the University of California, Berkeley, majoring in music. He became fascinated by the challenge of understanding and modeling human behavior only after he took a required economics class. He realized, “All of the biggest problems we face as a society, indeed as human beings, come down to research questions in the social sciences!”

Driven by his desire to influence public policy, Gaubatz went on to earn one master’s degree from the Fletcher School of Law and Diplomacy and another from Princeton Theological Seminary. He then earned his PhD in political science from Stanford University and, several prestigious fellowships later, is now on the faculty in the graduate program in international studies of the Department of Political Science & Geography at Old Dominion University. He describes his career in research as “a life of posing and answering questions, of trying to think about things in new and more interesting ways.”

Gaubatz’s advice for students interested in research careers focuses on the ongoing revolution in information technology:

We are in the middle of a revolution in data creation and computing power. Just 25 years ago, people could make a career from knowing information. A huge amount of information is now increasingly available to everyone who carries a phone. The critical skill is knowing how to build new ideas from the organization and analysis of that information, and being able to communicate those ideas effectively. Students need to focus on filling their toolboxes with those analytic and communication skills.

Variable-oriented research: Research that focuses attention on variables representing particular aspects of the cases studied and then examines the relations between these variables across sets of cases.

Research in the News

Can we Compare Equality on the Board Across the Globe?

Representation in boardrooms across the globe finds that women are sorely absent. The United States, which ranks fourth highest, found that women represent 16% of board members in Fortune 500 companies. Compare that with Norway, which enacted a law to ensure equality and in which 40% of public companies are directed by women. The inherently difficult obstacle in comparative analysis is that each country has a different measurement standard for gender equality.

1. What other measures would you suggest be considered for an international comparative study of gender equality?
2. In what ways could the utility of measures of gender equality be affected by variation in cultures?

News Source: Davidoff, Steven M. 2012. Seeking critical mass of gender equality in the boardroom. *New York Times*, September 12:B5.

Longitudinal Comparative Research

Dietrich Rueschemeyer et al. (1992) used a comparative historical method, combining the approaches, to explain why some nations in Latin America (excluding Central America) developed democratic politics, whereas others became authoritarian or bureaucratic–authoritarian states. First, Rueschemeyer et al. developed a theoretical framework that gave key attention to the power of social classes, state (government) power, and the interaction between social classes and the government. The researchers then classified the political regimes in each nation over time (Exhibit 11.11). Next, they noted how each nation varied over time relative to the variables they had identified as potentially important for successful democratization.

Exhibit 11.11 Classification of Regimes Over Time

	Constitutional Oligarchic	Authoritarian; Traditional, Populist, Military, or Corporatist	Restricted Democrat	Fully Democratic	Bureaucratic–Authoritarian
Argentina	before 1912	1930–46 1951–55 1955–58 1962–63	1958–62 1963–66	1912–30 1946–51 1973–76 1983–90	1966–73 1976–83
Bolivia	before 1930	1930–52 1964–82	1982–90	1952–64	
Brazil	before 1930	1930–45	1945–64 1985–90		1964–85
Chile	before 1920	1924–32	1920–24 1932–70 1990	1970–73	1973–89
Colombia	before 1936	1949–58	1936–49 1958–90		
Ecuador	1916–25	before 1916 1925–48 1961–78	1948–61 1978–90		
Mexico Paraguay		up to 1990 up to 1990			
Peru		before 1930 1930–39 1948–56 1962–63 1968–80	1939–48 1956–62 1963–68	1980–90	
Uruguay		before 1903 1933–42	1903–19	1919–33 1942–73 1984–90	1973–84
Venezuela		before 1935 1935–45	1958–68	1945–48 1968–90	

Source: Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. Capitalist development and democracy. Chicago: University of Chicago Press. Used with permission.

Their analysis identified several conditions for initial democratization: consolidation of state power (ending overt challenges to state authority), expansion of the export economy (reducing conflicts over resources), industrialization (increasing the size and interaction of middle and working classes), and some agent of political articulation of the subordinate classes (which could be the state, political parties, or mass movements). Historical variation in these conditions was then examined in detail.

The great classical sociologists also used comparative methods, although their approach was less systematic. For example, Max Weber's (Bendix 1962: 268) comparative sociology of religions contrasted Protestantism in the West, Confucianism and Taoism in China, Hinduism and Buddhism in India, and Ancient Judaism. As Bendix (1962) explained,

> His [Weber's] aim was to delineate religious orientations that contrasted sharply with those of the West, because only then could he specify the features that were peculiar to Occidental [Western] religiosity and hence called for an explanation . . . to bring out the distinctive features of each historical phenomenon. (p. 268)

So, for example, Weber concluded that the rise of Protestantism, with its individualistic approach to faith and salvation, was an important factor in the development of capitalism.

Research That Matters

Is an increase in democratic freedoms in nations associated with greater representation of women in powerful political positions? Prior research indicates that this is not the case; in fact, case studies have shown a drop in women's representation in government in some countries that have adopted democratic forms of governance. However, there are many complicating factors in the histories of particular nations, including whether gender quotas were implemented and the nature of the prior regime. Kathleen Fallon, Liam Swiss, and Jocelyn Viterna conducted a historical comparative research project to investigate why more democracy can be associated with fewer women in government. Fallon, Swiss, and Viterna collected data from 118 developing countries over a 34-year period. The dependent variable in the analysis was the percentage of seats held by women in the national legislature or its equivalent. The researchers distinguished countries transitioning from civil strife, authoritarian regimes, and communist regimes, and they accounted for the use of quotas for women as well as the extent of democratic practices and the differences in national culture.

The results indicate that women's legislative representation drops after democratizing changes begin, but then increases with additional elections. However, the strength of this pattern varies with the type of pre-democratic regime and the use of quotas. The nature of the *process* of democratic change is critical to understanding its outcome for women.

Cautions for Comparative Analysis

Of course, ambitious methods that compare different countries face many complications. The features of the cases selected for comparison have a large impact on the researcher's ability to identify influences. Cases should be chosen for their difference of key factors hypothesized to influence the outcome of interest and their similarity on other, possibly confounding, factors (Skocpol 1984: 383). For example, to understand how industrialization influences democracy, you would need to select cases for comparison that differ in industrialization, so that you could then see if they differ in democratization (King, Keohane, & Verba 1994: 148–152). Nonetheless, relying on just a small number of cases for comparisons introduces uncertainty into the conclusions (de Vaus 2008: 256). The focus on comparisons between nations may itself be a mistake for some analyses. National boundaries often do not correspond to key cultural differences, so comparing subregions within countries or larger cultural units that span multiple countries may make more sense for some analyses (de Vaus 2008: 258). Comparing countries that have fractured along cultural or religious divides simply by average characteristics would obscure many important social phenomena.

Video Link

Watch a clip about current international comparative research.

With cautions such as these in mind, historical and comparative methods allow for rich descriptions of social and political processes in different nations or regions as well as for causal inferences that reflect a systematic, defensible weighing of the evidence. Data of increasingly good quality are available on a rapidly expanding number of nations, creating many opportunities for comparative research. We cannot expect one study comparing the histories of a few nations to control adequately for every plausible alternative causal influence, but repeated investigations can refine our understanding and lead to increasingly accurate causal conclusions (King et al. 1994: 33).

Ethical Issues in Unobtrusive Methods

Ethical concerns arise when using unobtrusive measures that involve observing people, analyzing pictures of them, or collecting evidence of their activities. Although the potential harm to research participants may be delayed, it can still occur unless care is used to avoid disclosing identities—including covering faces in photos that are published. Pictures of individuals engaging in activities in public settings do not create as many concerns, but even such pictures may reveal behaviors that the participants would not want to be disclosed.

Analysis of historical documents, documents from other countries, or content in media does not create the potential for harm to human subjects that can be a concern when collecting primary data. It is still important to be honest and responsible in working out arrangements for data access when data must be obtained from designated officials or data archivists, but many data are available easily in libraries or on the web. Researchers in the United States who conclude that they are being denied access to public records of the federal government may be able to obtain the data by filing a Freedom of Information Act (FOIA) request. The FOIA stipulates that all persons have a right to access all federal agency records unless the records are specifically exempted (Riedel 2000: 130–131). Researchers who review historical or government documents must also try to avoid embarrassing or otherwise harming named individuals or their descendants by disclosing sensitive information.

Ethical concerns are multiplied when surveys are conducted or other data are collected in other countries. If the outside researcher lacks much knowledge of local norms, values, and routine activities, the potential for inadvertently harming subjects is substantial. For this reason, cross-cultural researchers should spend time learning about each of the countries in which they plan to collect primary data and strike up collaborations with researchers in those countries (Hantrais & Mangen 1996). Local advisory groups may also be formed in each country so that a broader range of opinion is solicited when key decisions must be made. Such collaboration can also be invaluable when designing instruments, collecting data, and interpreting results.

Cross-cultural researchers who use data from other societies have a particular obligation to try to understand the culture and norms of those societies before

they begin secondary data analyses. It is a mistake to assume that questions asked in other languages or cultural contexts will have the same meaning as when asked in the researcher's own language and culture, so a careful, culturally sensitive process of review by knowledgeable experts must precede measurement decisions in these projects. Researchers must become familiar with gender norms in the societies they seek to study because they may result in cross-country variation in responses to survey questions, willingness to participate in surveys, definitions of terms used in government statistics such as labor participation, and distortions in statistical data (Ayhan 2001).

Conclusion

We've covered a huge range of research methods in this chapter, but all of them intervene relatively little in the lives of people they study, unlike participant observation, surveys, or interviews; in that sense all are "unobtrusive." Some of them represent among the finest examples of classical and contemporary social science, and are capable of addressing sweeping topics of international importance. Ideally, in your own research you can use and combine different methods, as a way of compensating for the weaknesses of each, to improve the validity of your findings. The creative methods we suggested at the beginning of this chapter should help with that—and perhaps be enjoyable to develop and use, as well.

Interactive Exercises Link

Historical and Comparative Methods

Key Terms

Archival data 262
Case-oriented research 272
Content analysis 265
Contrived observation 264
Event-structure analysis 273
Holistic research 272
Narrative explanation 272
Oral history 273
Physical traces 260
Reactive methods 259
Unobtrusive measures 259
Variable-oriented research 276

Highlights

- Many social science projects rely on methods such as surveys, interviews, or participant observations that are inherently reactive, in that they may change the behavior they are intended to study. Unobtrusive measures try to avoid this weakness in research.
- Unobtrusive measures can be based on physical traces, archives, or observations.
- Content analysis is a tool for systematic quantitative analysis of documents and other textual data. It requires careful testing and control of coding procedures to achieve reliable measures.
- The central insight behind historical and comparative methods is that we can improve our understanding of social processes when we make comparisons with other times and places.
- Event-structure analysis is a systematic qualitative approach to developing an idiographic causal explanation for a key event.
- Oral history provides a means of reconstructing past events. Data from other sources should be used whenever possible to evaluate the accuracy of memories.
- Comparative methods may be cross-sectional, such as when variation between country characteristics is compared, or longitudinal, in which developmental patterns are compared between countries.
- Analysis of historical documents, documents from other countries, or content in media usually creates less potential for harm to human subjects than when collecting primary data, but it is still important to be honest and responsible in working out arrangements for data access when data must be obtained from designated officials or data archivists. Unobtrusive measures obtained from physical traces or observations require attention to the ethical issues also relevant in qualitative research.

Student Study Site

Related Titles in Research Methods and Statistics From SAGE

Investigating the Social World: The Process and Practice of Research, Eighth Edition, by Russell K. Schutt

The Practice of Research in Social Work, Third Edition, by Rafael J. Engel and Russell K. Schutt

The Practice of Research in Criminology and Criminal Justice, Fifth Edition, by Ronet K. Bachman and Russell K. Schutt

Fundamentals of Research in Criminology and Criminal Justice, Third Edition, by Ronet K. Bachman and Russell K. Schutt

Designing Surveys: A Guide to Decisions and Procedures, Second Edition, by Ronald F. Czaja and Johnny Blair

A Guide to Qualitative Field Research, Second Edition, by Carol A. Bailey

Adventures in Social Research, Seventh Edition, by Earl Babbie, Fred Halley, William E. Wagner, III, and Jeanne Zaino

Adventures in Criminal Justice Research, Fourth Edition, by Kim Logio, George Dowdall, Earl Babbie, and Fred Halley

Social Statistics for a Diverse Society, Sixth Edition, by Chava Frankfort-Nachmias and Anna Leon-Guerrero

Essentials of Social Statistics for a Diverse Society by Anna Leon-Guerrero and Chava Frankfort-Nachmias

Research Methods in Psychology, Second Edition, by Paul G. Nestor and Russell K. Schutt

Research Methods in Education by Joseph Check and Russell K. Schutt

Fundamentals of Social Work Research, Second Edition, by Rafael J. Engel and Russell K. Schutt

Constructing Social Research by Charles C. Ragin and Lisa M. Amoroso

Using IBM® SPSS® Statistics for Social Statistics and Research Methods, Third Edition, by William E. Wagner, III

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore | Boston

The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. The creative measures suggested by Webb et al. as well as those described in the beginning of this chapter span a wide range of approaches. Can you think of other unobtrusive measures you might use?
2. Review the differences between case-oriented, historically specific, inductive explanations and those that are more variable oriented, theoretically general, and deductive. List several arguments for and against each approach. Which is more appealing to you and why?
3. What historical events have had a major influence on social patterns in the nation? The possible answers are too numerous to list, ranging from any of the wars to major internal political conflicts, economic booms and busts, scientific discoveries, and legal changes. Pick one such event in your own nation for this exercise. Find one historical book on this event, and list the sources of evidence used. What additional evidence would you suggest for a social science investigation of the event?
4. Susan Olzak, Suzanne Shanahan, and Elizabeth McEneaney (1996) developed a nomothetic causal explanation of variation in racial rioting in the United States over time, whereas Griffin's (1993) explanation of a lynching can be termed *idiographic.* Discuss the similarities and differences between these types of causal explanation. Use these two studies to illustrate the strengths and weaknesses of each.

Finding Research

1. Paul Ekman, the psychologist cited who studies evidence of emotions in people's faces, has written extensively on this topic, and his work is widely used by police departments and even intelligence agencies. Find and read his findings on how to spot if someone is lying.
2. The journals *Social Science History* and *Journal of Social History* report many studies of historical processes. Select one article from a recent journal issue about a historical process used to explain some event or other outcome. Summarize the author's explanation. Identify any features of the explanation that are temporal, holistic, and conjunctural. Prepare a chronology of the important historical events in that process. Do you agree with the author's causal conclusions? What additional evidence would strengthen the author's argument?

Critiquing Research

1. What would be the weaknesses of using graffiti, such as in the Klofas and Cutshall (1985) study, to determine what prison inmates are thinking about? Might there be other ways of gathering such information that could be more accurate? What would be their weaknesses?

Doing Research

1. If you've read some of Ekman's work as suggested in "Finding Research," use his methods to watch people at some event—a sporting competition, maybe, or a reception. Keep track in detail of what they look like, and see if you can spot unexpected or socially awkward reactions. What might they mean?
2. Consider the media that you pay attention to in your social world. How could you design a content analysis of the messages conveyed by these media? What research questions could you help to answer by adding a comparison with another region or country to this content analysis?
3. Select a major historical event or process, such as the Great Depression, World War II, the civil rights movement, or the war in Iraq. Why do you think this event happened? Now, select an historical or comparative method that you think could be used to test your explanation. Why did you choose this method? What type of evidence would support your proposed explanation? What problems might you face in using this method to test your explanation?
4. Using your library's government documents collection or the U.S. Census site on the web, select one report by the U.S. Census Bureau about the population of the United States or some segment of it. Outline the report and list all the tables included in it. Summarize the report in two paragraphs. Suggest a historical or comparative study for which this report would be useful.
5. Consider the comparative historical research by Rueschemeyer et al. (1992) on democratic politics in Latin America. What does comparison between nations add to the researcher's ability to develop causal explanations?
6. Exhibit 11.12 identifies voting procedures and the level of turnout in 1 election for 10 countries. Do voting procedures appear to influence turnout in these countries?

Exhibit 11.18 Voting Procedures in 10 Countries

	Voting Age	Number of Days Polling Booth Open	Voting Day on Work Day or Rest Day	Postal Voting	Proxy Voting	Constituency Transfer	Advance Voting	Voter Turnout (in %)	Year (P=presidential, L = legislative election)
Switzerland	20	2	Rest day	Automatic for armed forces, otherwise by application 4 days before voting	Varies by canton	No	No	46	1991L
Taiwan	20	1	Rest day					72	1992L
Thailand	20	1	Rest day	No				62	1995L
Turkey	20	1	Rest day	No	No	Special polling stations at border posts for citizens residing abroad	No	80	1991L
Ukraine	18	1	Rest day					71.6	1994P
United Kingdom	18	1	Work day	On application	On application	No	No	77.8	1992L
United States	18	1	Work day	By application; rules vary across states	In some states for blind and disabled	No		51.5	1992P
Uruguay	18	1	Rest day	No	No	No		89.4	1994P
Venezuela	18	1	Rest day	No	Assisted voting for blind and disabled	No	No	60	1993P
Zambia		1	Work day		No			50	1991P

Source: LeDuc, Lawrence, Richard G. Niemi, and Pippa Norris (Eds.). 1996. Comparing democracies: Elections and voting in global perspective. Thousand Oaks, CA: Sage, 19, Figure 1.3.

Ethics Questions

1. Facebook and other popular social media sites routinely collect, use, and sell massive amounts of personal data. Do you think that's ethically right? When could it be right, and when wrong? What about experimentation on users, such as giving some users certain information and others not? Do you think a blanket waiver, such as what all users must sign when joining many sites, provides a sufficient level of consent?
2. Oral historians can uncover disturbing facts about the past. What if a researcher were conducting an oral history project such as the Depression Writer's Project and learned from an interviewee about his previously undisclosed involvement in a predatory sex crime many years ago? Should the researcher report what he learned to a government attorney who might decide to bring criminal charges? What about informing the victim or her surviving relatives? Would it matter if the statute of limitations had expired, so that the offender could not be prosecuted any longer? Would it matter if the researcher were subpoenaed to testify before a grand jury?
3. In this chapter's ethics section, we recommended that researchers who conduct research in other cultures form an advisory group of local residents to provide insight into local customs and beliefs. What are some other possible benefits of such a group for cross-cultural researchers? What disadvantages might arise from use of such a group?

Video Interview Questions

1. Listen to the researcher interview for Chapter 11 at edge.sagepub.com/chamblissmssw5e.
2. What caused Cinzia Solari's research question to change? What was the comparative element in her research?
3. How did Solari build rapport between her and the migrant workers she was trying to research? Why is this step important when doing qualitative research?

Chapter 12 Evaluation Research

©iStockphoto.com/asiseeit

Learning Objectives

1. Describe the history of evaluation research and its current status.
2. Diagram the evaluation research process as a feedback system.
3. Present arguments for and against stakeholder-driven evaluation.
4. Explain the concept of "black box" evaluation and the value of opening the black box.
5. Discuss the role of program theory and its value in evaluation research.
6. Define the five primary types of program evaluation research, and explain when each is appropriate.
7. List two advantages of including multiple outcomes in an evaluation research project.
8. Write an argument supporting or opposing research to evaluate social programs.

Drug Abuse Resistance Education (D.A.R.E.), as you probably know, is offered in elementary schools across the United States. For parents worried about drug abuse among youth and for many concerned citizens, the program has immediate appeal. It brings a special police officer into the schools once a week to talk to students about the hazards of drug abuse and to establish a direct link between local law enforcement and young people. You only have to check out bumper stickers or attend a few Parent-Teacher Association (PTA) meetings to learn that it's a popular program. It is one way many local governments have implemented antidrug policies.

And it is appealing. D.A.R.E. seems to improve relations between the schools and law enforcement and to create a positive image of the police in the eyes of students.

> It's a very positive program for kids . . . a way for law enforcement to interact with children in a nonthreatening fashion. . . . D.A.R.E. sponsored a basketball game. The middle school jazz band played. . . . We had families there. . . . D.A.R.E. officers lead activities at the [middle school]. . . . Kids do woodworking and produce a play. (Taylor 1999: 1, 11)

For some, the positive police–community relationships created by the program are enough to justify its continuation (Birkeland, Murphy-Graham, & Weiss

2005: 248), but most communities are concerned with its value in reducing drug abuse among children. Does D.A.R.E. lessen the use of illicit drugs among D.A.R.E. students? Does it do so while they are enrolled in the program or, more important, after they enter middle or high school? Unfortunately, evaluations of D.A.R.E. using social science methods led to the conclusion that students who participated in D.A.R.E. were no less likely to use illicit drugs than were comparable students who did not participate in D.A.R.E. (Ringwalt et al. 1994; West & O'Neal 2004).

If, like us, you have a child who enjoyed D.A.R.E., or were yourself a D.A.R.E. student, this may seem like a depressing way to begin a chapter on evaluation research. Nonetheless, it drives home an important point: To know whether social programs work, or how they work, we have to evaluate them systematically and fairly, whether we personally like the programs or not. And there's actually an optimistic conclusion to this introductory story: Evaluation research can make a difference. After the accumulation of evidence that D.A.R.E. programs were ineffective (West & O'Neal 2004), a "new" D.A.R.E. program was designed that engaged students more actively (Toppo 2002).

> Gone is the old-style approach to prevention in which an officer stands behind a podium and lectures students in straight rows. New D.A.R.E. officers are trained as "coaches" to support kids who are using research-based refusal strategies in high-stakes peer-pressure environments. (D.A.R.E. 2008)

Of course, the "new D.A.R.E." is now being evaluated, too. Sorry to say, one early quasi-experimental evaluation in 17 urban schools, funded by D.A.R.E. America, found no effect of the program on students' substance use (Vincus et al. 2010).

In this chapter, you will read about a variety of social program evaluations, alternative approaches to evaluation, and the different types of evaluation research and review ethical concerns. You should finish the chapter with a much better understanding of how the methods of applied social research can help improve society.

What Is the History of Evaluation Research?

Evaluation research is not a method of data collection, like survey research or experiments; nor is it a unique component of research designs, like sampling or measurement. Instead, evaluation research is conducted for a distinctive purpose: to investigate social programs (such as substance abuse treatment programs, welfare programs, criminal justice programs, or employment and training programs). For each project, an evaluation researcher must select a research design and method of data collection that are useful for answering the particular research questions posed and appropriate for the particular program investigated.

Audio Link

Listen to a variety of podcasts on current trends and issues in evaluation research.

So, you can see why we placed this chapter after most of the others in the text. When you review or plan evaluation research, you have to think about the research process as a whole and how different parts of that process can best be combined.

The development of evaluation research as a major enterprise followed on the heels of the expansion of the federal government during the Great Depression and World War II. Large Depression-era government outlays for social programs stimulated interest in monitoring program output, and the military effort in World War II led to some of the necessary review and contracting procedures for sponsoring evaluation research. However, not until the Great Society programs of the 1960s did evaluation begin to be required when new social programs were funded (Dentler 2002; Rossi & Freeman 1989: 34). The World Bank and International Monetary Fund (IMF) began to require evaluation of the programs they fund in other countries (Dentler 2002: 147). More than 100 contract research and development firms began in the United States between 1965 and 1975, and many federal agencies developed their own research units. The RAND Corporation expanded from its role as a U.S. Air Force planning unit into a major social research firm; SRI International spun off from Stanford University as a private firm; and Abt Associates in Cambridge, Massachusetts, which began

in a garage in 1965, grew to employ more than 1,000 employees in five offices in the United States, Canada, and Europe.

With the decline of many Great Society programs in the early 1980s, many such evaluation research firms closed down. But recently, with more calls for government "accountability," the evaluation research enterprise has been growing again. The Community Mental Health Act Amendments of 1975 (Public Law 94–63) required quality assurance (QA) reviews, which often involve evaluation-like activities (Patton 2002: 147–151). The Government Performance and Results Act of 1993 required some type of evaluation of all government programs (Office of Management and Budget n.d.). At century's end, the federal government was spending about $200 million annually on evaluating $400 billion in domestic programs, and the 30 major federal agencies had between them 200 distinct evaluation units (Boruch 1997). In 1999, the new Governmental Accounting Standards Board urged that more attention be given to "service efforts and accomplishments" in standard government fiscal reports (Campbell 2002).

The growth of evaluation research is also reflected in the social science community. The American Evaluation Association was founded in 1986 as a professional organization for evaluation researchers (merging two previous associations) and is the publisher of an evaluation research journal. In 1999, evaluation researchers founded the Campbell Collaboration to publicize and encourage systematic review of evaluation research studies. Their online archive contains 10,449 reports on randomized evaluation studies (Davies, Petrosino, & Chalmers 1999).

Careers and Research

Mary Anne Casey, PhD, Consultant

Source: Mary Anne Casey

Mary Anne Casey sailed through her undergraduate work without any exposure to social research. Her career in research and evaluation was never part of a “grand plan”; she just

happened into it because of an assistantship in graduate school at the University of Minnesota. This graduate school experience—evaluating a regional foundation—fed her curiosity in research and evaluation.

After receiving her PhD, Casey worked for the State of Minnesota and the W. K. Kellogg Foundation and then joined a consulting firm. She weaves the lessons she has learned about research into her work, her writing on focus group interviewing (and a book with Richard Krueger on focus groups published by SAGE Publications), and her teaching at the University of Minnesota, University of South Florida, and University of Michigan. Throughout her career, she has never stopped learning.

Each study is an opportunity to learn. I've learned about vexing issues and I've learned strategies that make me a better interviewer and analyst. The greatest reward is the honor of listening to people from a variety of backgrounds on intriguing topics: Midwest farmers on corn rootworms, veterans on their mental health care, mothers of new babies on home health care visits, teenagers on birth control, smokers on quitting, community members on garbage pickup, faculty on job satisfaction, and kids on what would get them to eat more fruits and vegetables. As a result, I know that there are multiple ways to see any issue. I believe this has made me less judgmental.

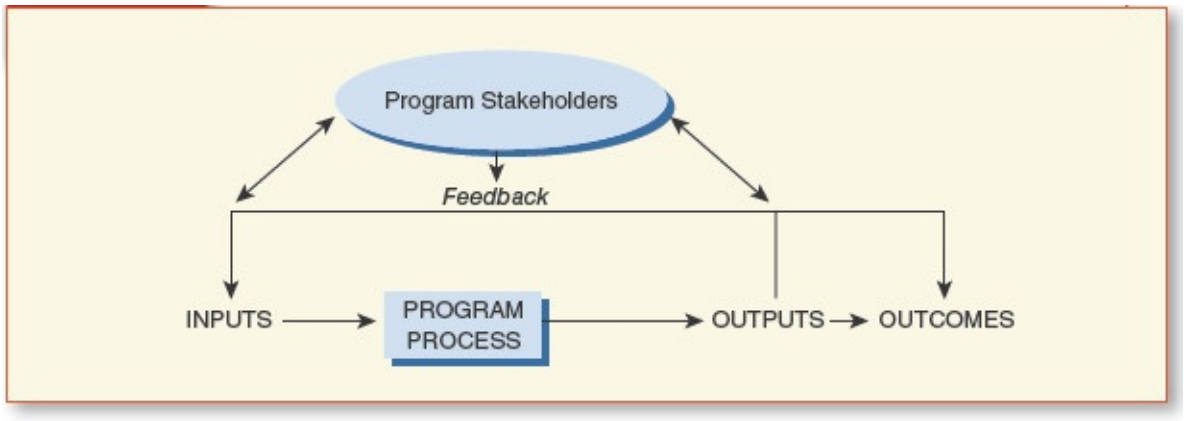
Casey relishes analysis and finding just the right way to convey what people have shared. She urges students interested in research careers to hone their skills as listeners.

I hope my writing and teaching about focus group interviewing convinces others that careful listening is valuable and doable. We need good listeners.

What Is Evaluation Research?

Exhibit 12.1 illustrates the process of evaluation research as a simple systems model. First, clients, customers, students, or some other persons or units—cases—enter the program as **inputs**. (Notice that this model regards programs as machines, with clients—people—seen as raw materials to be processed.) Students may begin a new school program, welfare recipients may enroll in a new job-training program, or crime victims may be sent to a victim advocate. Resources and staff required by a program are also program inputs.

Exhibit 12.1 A Model of Evaluation



Source: Adapted from Martin, Lawrence L., and Peter M. Kettner. 1996. Measuring the performance of human service programs. Thousand Oaks, CA: Sage.

Next, some service or treatment is provided to the cases. This may be attendance in a class, assistance with a health problem, residence in new housing, or receipt of special cash benefits. This process of service delivery—the **program process**—may be simple or complicated, short or long, but it is designed to have some impact on the cases as inputs are consumed and outputs are produced.

Program **outputs** are the direct product of the program's service delivery process. They could include clients served, case managers trained, food parcels delivered, or arrests made. The program outputs may be desirable in themselves, but primarily they indicate that the program is operating.

Program **outcomes** indicate the impact of the program on the cases that have been processed. Outcomes can range from improved test scores or higher rates of job retention to fewer criminal offenses and lower rates of poverty. There are likely to be multiple outcomes of any social program, some intended and some unintended, some viewed as positive and others viewed as negative.

Through a **feedback** process, variation in outputs and outcomes can influence the inputs to the program. If not enough clients are being served, recruitment of new clients may increase. If too many negative side effects result from a trial medication, the trials may be limited or terminated. If a program does not lead to improved outcomes, clients may be sent elsewhere.

Evaluation research itself is really just a systematic approach to feedback; it strengthens the feedback loop through credible analyses of program operations and outcomes. Evaluation research also broadens this loop to include connections to parties outside of the program itself. A funding agency or political authority may mandate the research, outside experts may be brought in to conduct the research, and the evaluation research findings may be released to the public, or at least to funders, in a formal report.

Researcher Interview Link

Watch a researcher describe a specific study that used evaluation research.

The evaluation process as a whole, and the feedback in particular, can be understood only in relation to the interests and perspectives of program stakeholders. **Stakeholders** are those individuals and groups who have some basis of concern for the program. They might be clients, staff, managers, funders, or the public. The board of a program or agency, the parents or spouses of clients, the foundations that award program grants, the auditors who monitor program spending, the members of Congress—each is a potential program stakeholder, and each has an interest in the outcome of any program evaluation. Some may fund the evaluation, some may provide research data, and some may review—or even approve—the research report (Martin & Kettner 1996: 3). Who the program stakeholders are, and what role they play in the program evaluation, can have tremendous consequences for the research.

Thus, there are real differences between traditional social science and evaluation

research (Posavac & Carey 1997). Social science is motivated by theoretical concerns and is guided by the standards of research methods without consideration (ideally) for political factors. It examines specific organizations for what, in general, we can learn from them, not for improving that one organization. Practical ramifications, for particular programs, are not usually of any import. For evaluation research, however, the particular program and its impact are paramount. How the program works also matters—not to advance a theory but to improve the program. Finally, stakeholders of all sorts—not an abstract "scientific community"—have a legitimate role in setting the research agenda and may well intervene, even when they aren't supposed to. But overall, there is no sharp boundary between the two approaches: In their attempt to explain how and why the program has an impact and whether the program is needed, evaluation researchers often bring social theories into their projects—but for immediately practical aims.

Inputs: Resources, raw materials, clients, and staff that go into a program.

Program process: The complete treatment or service delivered by the program.

Outputs: The services delivered or new products produced by the program process.

Outcomes: The impact of the program process on the cases processed.

Feedback: Information about service delivery system outputs, outcomes, or operations that is available to any program inputs.

Stakeholders: Individuals and groups who have some basis of concern with the program.

What Are the Alternatives in Evaluation Designs?

Evaluation research tries to learn if, and how, real-world programs produce results. But that simple statement covers a number of important alternatives in research design, including the following:

- *Black box or program theory*—Do we care how the program gets results?
- *Researcher or stakeholder orientation*—Whose goals matter most?
- *Quantitative or qualitative methods*—Which methods provide the best answers?
- *Simple or complex outcomes*—How complicated should the findings be?

Black Box or Program Theory

Most evaluation research tries to determine whether a program has the intended effect. If the effect occurred, the program "worked"; if the effect didn't occur, then, some would say, the program should be abandoned or redesigned. In this simple approach, the process by which a program produces outcomes is often treated as a "black box" in which the inside of the program is unknown. The focus of such research is whether cases have changed as a result of their exposure to the program between the time they entered as inputs and when they exited as outputs (Chen 1990). The assumption is that program evaluation requires only the test of a simple input/output model, like that in Exhibit 12.1. There may be no attempt to "open the black box" of the program process.

But there are good reasons to open the black box and investigate how the process works (or doesn't work). Consider recent research on welfare-to-work programs. The Manpower Demonstration Research Corporation reviewed findings from research on these programs in Florida, Minnesota, and Canada (Lewin 2001a). In each location, adolescents with parents in a welfare-to-work program were compared with a control group of teenagers whose parents were also on welfare but were *not* enrolled in welfare-to-work. In all three locations, teenagers in the welfare-to-work program families did *worse* in school than those in the control group.

But why did requiring welfare mothers to get jobs hurt their children's schoolwork? Unfortunately, because the researchers had not investigated program process—had not opened the black box—we can't know for sure. Martha Zaslow, an author of the resulting research report, speculated (as cited in Lewin 2001a) that

> parents in the programs might have less time and energy to monitor their adolescents' behavior once they were employed. . . . Under the stress of working, they might adopt harsher parenting styles. . . . The adolescents' assuming more responsibilities at home when parents got jobs was creating too great a burden. (p. A16)

Unfortunately, as Ms. Zaslow (as cited in Lewin 2001a) admitted, "We don't

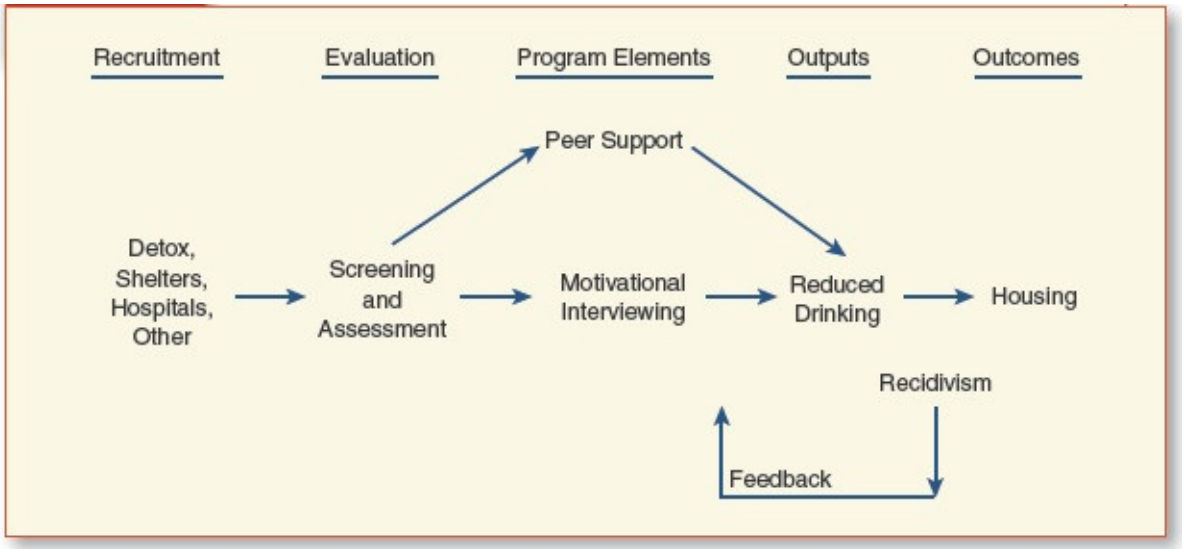
know exactly what's causing these effects, so it's really hard to say, at this point, what will be the long-term effects on these kids" (p. A16).

If an investigation of program process had been conducted, though, a **program theory** could have been developed. A program theory describes what has been learned about how the program has its effect. When a researcher has sufficient knowledge before the investigation begins, outlining a program theory can help to guide the investigation of program process in the most productive directions. This is termed a **theory-driven evaluation.**

A program theory specifies how the program is expected to operate and identifies which program elements are operational (Chen 1990: 32). In addition, a program theory specifies how a program is to produce its effects, thus improving the understanding of the relationship between the independent variable (the program) and the dependent variable (the outcome or outcomes). For example, Exhibit 12.2 illustrates the theory for an alcoholism treatment program. It shows that persons entering the program are expected to respond to the combination of motivational interviewing and peer support. A program theory also can decrease the risk of failure when the program is transported to other settings because it will help to identify the conditions required for the program to have its intended effect.

Program theory can be either descriptive or prescriptive (Chen 1990). *Descriptive theory* specifies impacts that are generated and how this occurs. It suggests a causal mechanism, including intervening factors and the necessary context for the effects. Descriptive theories are generally empirically based. *Prescriptive theory* specifies what *ought* to be done by the program and is not actually tested. Prescriptive theory specifies how to design or implement the treatment, what outcomes should be expected, and how performance should be judged. Comparison of the program's descriptive and prescriptive theories can help to identify implementation difficulties and incorrect understandings that can be fixed (Patton 2002: 162–164).

Exhibit 12.2 The Program Theory for a Treatment Program for Homeless Alcoholics



Program theory: A descriptive or prescriptive model of how a program operates and produces effects.

Theory-driven evaluation: A program evaluation guided by a theory that specifies the process by which the program has an effect.

Researcher or Stakeholder Orientation

Whose prescriptions direct the program? What outcomes it should achieve? Whom it should serve? Most social science assumes that the researcher decides. Research results are usually reported in professional journals or conferences, where scientific standards determine how it is judged. In program evaluation, however, the program sponsors or a government agency often set the research question; in consulting projects for businesses, the client—a manager, perhaps, or a division president—decides what question researchers will study. Research findings are reported to these authorities, who most often also specify the outcomes to be investigated. The primary evaluator of evaluation research, then, is the funding agency, not the professional social science community. Evaluation research is research for a client, and its results may directly affect the services, treatments, or even punishments (in the case of prison studies, for example) that program users receive. Who pays the piper, picks the tune.

Should the evaluation researcher insist on designing the project and specifying its goals? Or should she accept the suggestions and goals of the funding agency? What role should program staff and clients play? What responsibility does the researcher have to politicians and taxpayers when evaluating government-funded programs?

Various evaluation researchers have answered these questions through different —stakeholder, social science, and integrative—approaches (Chen 1990: 66–68). **Stakeholder approaches** encourage researchers to be responsive to program stakeholders. Issues for study are to be based on the views of people involved with the program, and reports are to be made to program participants (Stake 1975). The researcher develops the program theory to clarify and develop the key stakeholders' theory of the program (Wholey 1987). In one stakeholder approach, termed *utilization-focused evaluation*, the evaluator forms a task force of program stakeholders who help to shape the evaluation project so that they are most likely to use its results (Patton 2002: 171–175). In evaluation research termed *action research* or *participatory research*, program participants are engaged with the researchers as coresearchers and help design, conduct, and report the research. One research approach, termed *appreciative inquiry*, eliminates the professional researcher altogether in favor of a structured dialogue about needed changes among program participants themselves (Patton 2002:

177–185).

Egon Guba and Yvonna Lincoln (1989) argue for a stakeholder approach in their book, *Fourth Generation Evaluation:*

> The stakeholders and others who may be drawn into the evaluation are welcomed as equal partners in every aspect of design, implementation, interpretation, and resulting action of an evaluation—that is, they are accorded a full measure of political parity and control . . . determining what questions are to be asked and what information is to be collected on the basis of stakeholder inputs. (p. 11)

Social science approaches, in contrast, emphasize researcher expertise autonomy to develop the most trustworthy, unbiased program evaluation. These approaches assume that "evaluators cannot passively accept the values and views of the other stakeholders" (Chen 1990: 78). Instead, the researcher derives a program theory from information on how the program operates and current social science theory, not from the views of stakeholders. In one somewhat extreme form of this approach, *goal-free evaluation*, researchers do not even permit themselves to learn what goals the program stakeholders have for the program. Instead, the researcher assesses and then compares the needs of participants to a wide array of program outcomes (Scriven 1972). The goal-free evaluator wants to see the unanticipated outcomes and to remove any biases caused by knowing the program goals in advance.

Of course, there are disadvantages to both stakeholder and social science approaches to program evaluation. If stakeholders are ignored, researchers may find that participants are uncooperative, that their reports are unused, and that the next project remains unfunded. If social science procedures are neglected, standards of evidence will be compromised, conclusions about program effects will likely be invalid, and results are unlikely to be generalizable to other settings. These equally undesirable possibilities have led to several attempts to develop more integrated approaches to evaluation research.

Integrative approaches attempt to cover issues of concern to both stakeholders and evaluators (Chen & Rossi 1987: 101–102). The emphasis given to either stakeholder or scientific concerns varies with the specific circumstances. Integrative approaches seek to balance responsiveness to stakeholders with

objectivity and scientific validity. Evaluators negotiate regularly with key stakeholders during the planning of the research; preliminary findings are reported back to decision makers so they can make improvements; and when the final evaluation is conducted, the research team may operate more autonomously, minimizing intrusions from program stakeholders. Evaluators and clients thus work together.

Stakeholder approaches (to evaluation): An orientation to evaluation research that expects researchers to be responsive primarily to the people involved with the program.

Social science approaches (to evaluation): An orientation to evaluation research that expects researchers to emphasize the importance of researcher expertise and maintenance of autonomy from program stakeholders.

Integrative approaches (to evaluation): An orientation to evaluation research that expects researchers to respond to the concerns of people involved with the program stakeholders, as well as to the standards and goals of the social scientific community.

Quantitative or Qualitative Methods

Quantitative and qualitative approaches to evaluation each have their strengths and appropriate uses. Quantitative research, with its clear percentages and numerical scores, allows quick comparisons over time and categories and, thus, is typically used in attempts to identify the effects of a social program. With numbers, you can systematically track change over time or compare outcomes between an experimental and a control group. Did the response times of emergency personnel tend to decrease? Did the students' test scores increase more in the experimental group than in the control group? Did housing retention improve for all subjects or just for those who were not substance abusers? Quantified results also can prevent distraction by the powerful anecdote, forcing you to see what happens in most cases, not just in the dramatic cases; they "force you to face reality," as a friend of ours puts it.

Qualitative methods, however, can add depth, detail, and nuance; they can clarify the meaning of survey responses and reveal more complex emotions and judgments people may have (Patton 2002). Perhaps the greatest contribution qualitative methods can make is in investigating program process—finding out what is "inside the black box." Quantitative measures, like staff contact hours or frequency of complaints, can track items such as service delivery, but finding out how clients experience the program is best accomplished by directly observing program activities and interviewing staff and clients intensively.

For example, Timothy Diamond's (1992: 17) observational study of work in a nursing home shows how the somewhat cool professionalism of new program aides was softened to include a greater sensitivity to interpersonal relations:

Journal Link

Read an article about a mixed-methods approach to evaluating classroom dynamics.

> The tensions generated by the introductory lecture and. . . ideas of career professionalism were reflected in our conversations as we waited for the

> second class to get under way. Yet within the next half hour they seemed to dissolve. Mrs. Bonderoid, our teacher, saw to that. . . . "What this [work] is going to take," she instructed, "is a lot of mother's wit." "Mother's wit," she said, not "mother wit," which connotes native intelligence irrespective of gender. She was talking about maternal feelings and skills.

Surveys could have asked the aides how satisfied they were with their training but would not have revealed the subtler side of "mother's wit."

Qualitative methods also can uncover how different individuals react to the treatment. For example, a quantitative evaluation of student reactions to an adult basic skills program for new immigrants relied heavily on the students' initial statements of their goals. However, qualitative interviews revealed that most new immigrants lacked sufficient experience in America to set meaningful goals; their initial goal statements simply reflected their eagerness to agree with their counselors' suggestions (Patton 2002: 177–181).

Qualitative methods can, in general, help in understanding how social programs actually operate. In complex social programs, it is not always clear whether any particular features are responsible for the program's effect (or noneffect). Lisbeth B. Schorr, director of the Harvard Project on Effective Interventions, and Daniel Yankelovich, president of Public Agenda, put it this way: "Social programs are sprawling efforts with multiple components requiring constant mid-course corrections, the involvement of committed human beings, and flexible adaptation to local circumstances" (Schorr & Yankelovich 2000: A14). Schorr and Yankelovich pointed to the Ten Point Coalition, an alliance of black ministers that helped reduce gang warfare in Boston through multiple initiatives, "ranging from neighborhood probation patrols to safe havens for recreation" (p. A14). Qualitative methods help describe a complex, multifaceted program like this. In general, the more complex the social program, the more value that qualitative methods can add to the evaluation process.

Simple or Complex Outcomes

Few programs have only one outcome. Colleges provide not only academic education, for instance, but also—importantly—an amazingly efficient marketplace for potential spouses and lifetime friends. D.A.R.E. programs may not reduce drug use, but they often seem to improve student–police relations. Some outcomes are direct and intended; others happen only over time, are uncertain, and may well not be desired. A decision to focus exclusively on a single outcome—probably the officially intended one—can easily cause a researcher to ignore even more important results.

Sometimes a single policy outcome is sought but is found not to be sufficient, either methodologically or substantively. When Lawrence Sherman and Richard Berk (1984) evaluated the impact of an immediate arrest policy in cases of domestic violence in Minneapolis, they focused on recidivism—repeating the offense—as the key outcome. Similarly, the reduction of recidivism was the single desired outcome of the prison boot camps that began opening in the 1990s. Boot camps were military-style programs for prison inmates that provided tough, highly regimented activities and harsh punishment for disciplinary infractions with the goal of scaring inmates "straight." But these single-purpose programs, both designed to reduce recidivism, turned out not to be quite so simple to evaluate. The Minneapolis researchers found that there were no adequate single sources for recidivism in domestic violence cases, so they had to hunt for evidence from court and police records, perform follow-up interviews with victims, and review family member reports. More easily measured variables, such as partners' ratings of the accused's subsequent behavior, received more attention. Boot camp research soon concluded that the experience did not reduce recidivism, but some participants felt that boot camps did have some beneficial effects:

Audio Link

Listen to the results of a program with different outcomes.

> [A staff member] saw things unfold that he had never witnessed among

> inmates and their caretakers. . . . Profoundly affected the drill instructors and their charges. . . . Graduation ceremonies routinely reduced inmates . . . sometimes even supervisors to tears. . . . Here, it was a totally different experience. (Latour 2002: B7)

Some now argue that the failure of boot camps to reduce recidivism was caused by the lack of postprison support rather than to failure of the camps to promote positive change in inmates. Looking at recidivism rates alone would ignore some important positive results.

So despite the difficulties, most evaluation researchers attempt to measure multiple outcomes (Mohr 1992). One such evaluation appears in Exhibit 12.3. Project New Hope was an ambitious experimental evaluation of the impact of guaranteeing jobs to poor people (DeParle 1999). It was designed to answer the following question: If low-income adults are given a job at a sufficient wage, above the poverty level, with child care and health care assured, how many would ultimately prosper?

Exhibit 12.3 Outcomes in Project New Hope

Income and Employment (2nd program year)	New Hope	Control Group
Earnings	$6,602	$6,129
Wage subsidies	1,477	862
Welfare income	1,716	1,690
Food stamp income	1,418	1,242
Total income	11,213	9,915
% above poverty level	27%	19%
% continuously unemployed for 2 years	6%	13%
Hardships and Stress	**New Hope**	**Control Group**
% reporting:		
Unmet medical needs	17%	23%
Unmet dental needs	27%	34%
Periods without health insurance	49%	61%
Living in overcrowded conditions	14%	15%
Stressed much or all of the time	45%	50%
Satisfied or very satisfied with standard of living	65%	67%

Source: Adapted from DeParle, Jason. 1999. Project to rescue needy stumbles against the persistence of poverty. New York Times, May 15; A1, A10; Bos, J. M., and Manpower Demonstration Research Corporation. 1999. New hope for people with low incomes: Two-year results of a program to reduce poverty and reform welfare. New York: Manpower Demonstration Research Corp.

In Project New Hope, 677 low-income adults in Milwaukee, Wisconsin, were offered a job involving work for 30 hours a week, as well as child care and health care benefits. A control group did not receive the guaranteed jobs. The outcome? Only 27% of the 677 stuck with the job long enough to lift themselves out of poverty, and their earnings as a whole were only slightly higher than those of the control group. Levels of depression were not decreased, nor was self-

esteem increased by the job guarantee. But there were some positive effects: The number of people who never worked at all declined, and rates of health insurance and use of formal child care increased. Perhaps most important, the classroom performance and educational hopes of participants' male children increased, with the boys' test scores rising by the equivalent of 100 points on the SAT and their teachers ranking them as better behaved.

So did the New Hope program "work"? Clearly it didn't live up to initial expectations, but it certainly showed that social interventions can have some benefits. Would the boys' gains continue through adolescence? Longer-term outcomes would be needed. Why didn't girls (who were already performing better than the boys) benefit from their parents' enrollment in New Hope just as the boys did? A process analysis would add a great deal to the evaluation design. Collection of multiple outcomes, then, gives a better picture of program impact.

What Can an Evaluation Study Focus On?

Evaluation projects can focus on a variety of different questions related to social programs and their impact:

Research|Social Impact Link

Read more about evaluation studies.

- What is the level of need for the program?
- Can the program be evaluated?
- How does the program operate?
- What is the program's impact?
- How efficient is the program?

The question asked will determine what research methods are used.

Video Link

Watch a lecture on the elements of a needs assessment.

Needs Assessment

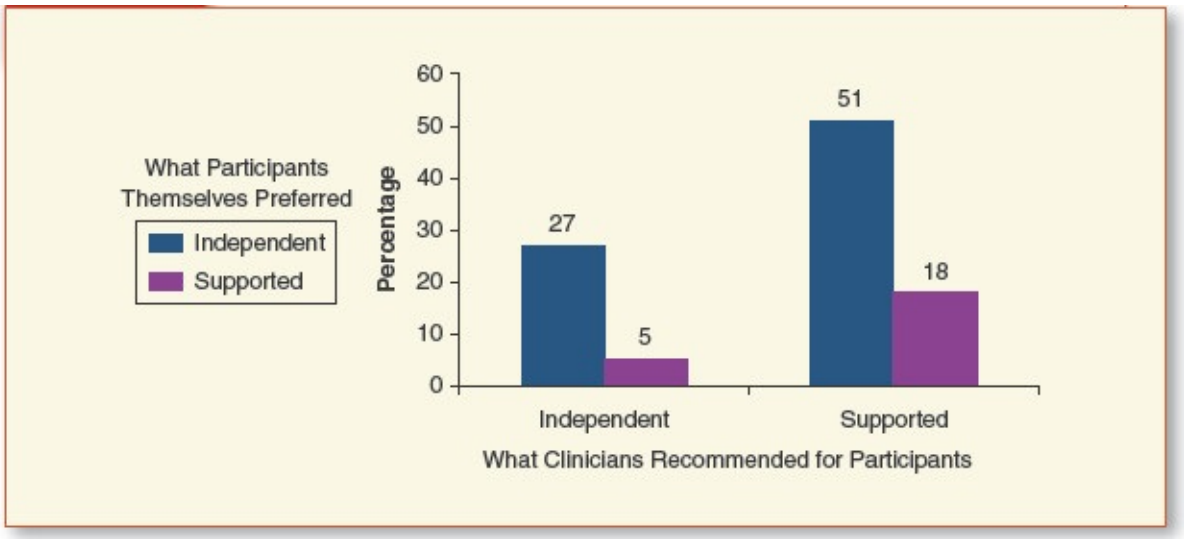
A **needs assessment** attempts, with systematic, credible evidence, to evaluate what needs exist in a population. Need may be assessed by social indicators, such as the poverty rate or the level of home ownership; interviews with local experts, such as school board members or team captains; surveys of populations potentially in need; or focus groups with community residents (Rossi & Freeman 1989).

It is not as easy as it sounds (Posavac & Carey 1997). Whose definitions of need should be used? How will we deal with ignorance of need? How can we understand the level of need without understanding the social context? (Short answer to that one: We can't!) What, after all, does *need* mean in the abstract?

The results of the Boston McKinney Project reveal the importance of taking a multidimensional approach to the investigation of need. The Boston McKinney Project evaluated the merits of providing formerly homeless mentally ill persons with staffed group housing as compared with individual housing (Schutt 2011). In a sense, you can think of the whole experiment as involving an attempt to answer the question "What type of housing do these persons 'need'?" Russ Schutt and his colleagues first examined this question at the start of the project, by asking each project participant which type of housing he or she wanted (Schutt & Goldfinger 1996) and by independently asking two clinicians to estimate which of the two housing alternatives would be best for each participant (Goldfinger & Schutt 1996).

Exhibit 12.4 displays the findings. The clinicians recommended staffed group housing for 69% of the participants (51 + 18), whereas most of the participants (78%) sought individual housing (27 + 51). In fact, there was no correspondence between the housing recommendations of the clinicians and the housing preferences of the participants (who did not know what the clinicians had recommended for them). So which perspective reveals the level of need for staffed group housing as opposed to individual housing?

Exhibit 12.4 Type of Residence: Preferred and Recommended



Source: Based on Goldfinger, Stephen M., and Russell K. Schutt. 1996. Comparisons of clinicians' housing recommendations of homeless mentally ill persons. Psychiatric Services 47(4): 413–415.

Of course, there's no objective answer. Policy makers' values, and their understanding of mental illness and homelessness, will influence which answer they prefer.

In general, it is a good idea to use multiple indicators of need. There is no absolute definition of need in this situation, nor is there in most projects. A good evaluation researcher will try to capture different perspectives on need and then help others make sense of the results.

Needs assessment: A type of evaluation research that attempts to determine the needs of some population that might be met with a social program.

Research in the News

What Motivates Policy Shifts?

A report from the American Civil Liberties Union (ACLU) documented a national corrections overhaul movement that is fragile but growing. After decades of increased incarceration and skyrocketing prison populations, states from the South to the Midwest are starting to see decreases. Much of the motivations behind these policy shifts are financial: housing prisoners is far more expensive than parole or drug treatment. Social science researchers have identified the value of alternative approaches to incarceration in many studies, but budget pressures are often more powerful than statistics.

1. What hypothesis would you propose to test about the value of reducing rates of incarceration, and what research design would you suggest using to test it?
2. Describe a possible research project about state incarceration policies using the policy research approach described in this chapter.

News Source: Savage, Charlie. 2011. Trend to lighten harsh sentences catches on in conservative states. *New York Times*, August 13:A12.

Evaluability Assessment

Evaluation research is pointless if the program cannot be evaluated. Yes, some type of study is always possible, but to identify specifically the effects of a program may not be possible within the available time and resources. So researchers may conduct an **evaluability assessment** to learn this in advance, rather than expend time and effort on a fruitless project (Patton 2002: 164).

Why might a social program not be evaluable?

- Management only wants to have its superior performance confirmed and does not really care whether the program is having its intended effects. This is a very common problem.
- Staff are so alienated from the agency that they don't trust any attempt sponsored by management to check on their performance.
- Program personnel are just "helping people" or "putting in time" without any clear sense of what the program is trying to achieve.
- The program is not clearly distinct from other services delivered by the agency and so can't be evaluated by itself.

Because they are preliminary studies to "check things out," evaluability assessments often rely on qualitative methods. Program managers and key staff may be interviewed, or program sponsors may be asked about the importance they attach to different goals.

Sometimes an evaluability assessment can actually help to solve problems. Discussion with program managers and staff can result in changes in program operations. The evaluators may use the evaluability assessment to sensitize participants to the importance of clarifying their goals and objectives. The knowledge gained can be used to refine evaluation plans.

Journal Link

Read a program evaluation that assessed social outcomes of older people.

The President's Family Justice Center (FJC) Initiative was initiated in President George W. Bush's administration to plan and implement comprehensive

domestic violence services that would provide “one stop shopping” for victims in need of services. In 2004, the National Institute of Justice contracted with Abt Associates in Cambridge, Massachusetts, to assess the evaluability of 15 pilot service programs that had been awarded a total of $20 million and to develop an evaluation plan. In September 2005, Abt researchers Meg Townsend, Dana Hunt, and William Rhodes reported on their evaluability assessment.

Abt’s assessment began with conversations to collect background information and perceptions of program goals and objectives from those who had designed the program. These conversations were followed by a review of the grant applications submitted by each of the 15 sites and phone conversations with site representatives. Site-specific data collection focused on the project’s history at the site, its stage of implementation, staffing plans and target population, program activities and stability, goals identified by the site’s director, apparent contradictions between goals and activities, and the state of data systems that could be used in the evaluation. Exhibit 12.5 shows the resulting logic model that illustrates the intended activities, outcomes, and impacts for the Alameda County, California, program. Although they had been able to begin the evaluability assessment process, Townsend and colleagues concluded that in the summer of 2005, none of the 15 sites were far enough along with their programs to complete the assessment.

Evaluability assessment: A type of evaluation research conducted to determine whether it is feasible to evaluate a program’s effects within the available time and resources.

Process Evaluation

What actually happens in a social program? In the New Jersey Income Maintenance Experiment, some welfare recipients received higher payments than others did (Kershaw & Fair 1976): simple enough, and not too difficult to verify that the right people received the intended treatment. In the Minneapolis experiment on the police response to domestic violence (Sherman & Berk 1984), some individuals accused of assaulting their spouses were arrested, whereas others were just warned. This is a little bit more complicated because the severity of the warning might have varied among police officers and, to minimize the risk of repeat harm, police officers were allowed to override the experimental assignment. To identify this deviation from the experimental design, the researchers would have had to keep track of the treatments delivered to each accused spouse and collect information on what officers actually *did* when they warned an accused spouse. This would be **process evaluation**—research to investigate the process of service delivery.

Exhibit 12.5 Alameda Family Justice Center Logic Model

Inputs	Activities	Outcomes	Impacts	Goals
• On-site partners • Intake systems • Client management process • Space design • Site location	**FJC** • Case management • Assistance with restraining orders • Assistance with police reports • Legal assistance • Advocacy • Medical care • Forensic exams • Assessments and referral for treatment • Counseling • Safety planning • Emergency food/cash/ transportation • Referral for shelter and other ongoing care • Assistance with public assistance • 24-hour helpline • Parenting classes • Child care • Rape crises services • Faith-based services • Job training • Translation services	**Victims** • Increase likelihood to access services • Increase demand for services • Increase usage of services • Increase frequency of cross-referrals or use of multiple services	**Victims** • Reduce tendency to blame oneself for abuse • Reduce conditions that prevent women from leaving • Increase likelihood of reporting incident • Increase likelihood of request for temporary/ permanent restraining orders • Increase likelihood of participating in prosecution	• Decrease incidents of DV • Decreased repeat victimizations • Decreased seriousness • Hold offenders accountable • Decrease repeat offenders • Break cycle of violence
	Community • Early intervention and prevention programming • FJC informational materials	**Community** • Increase knowledge of DV/SA/Elder Abuse • Increase awareness of services available	**Community** • Increase awareness of FJC • Decrease social tolerance for VAW*	

	Systems • Collaboration between government and nongov't providers • Improve access to batterer information	**Systems** • Improve DV policies and procedures • Increase understanding of each other's services • Increase coordination of services	**Systems** • Improve institutional response to DV • Decrease secondary trauma • Increase assurance of victim safety • Increase the number of successful criminal legal actions • Increase the number of successful civil legal actions	

*Violence Against Women

Source: Chen, Huey-Tsyh. 1990. Theory-driven evaluations. Newbury Park, CA: Sage, 210. Reprinted with permission from SAGE Publications, Inc.

Journal Link

Read an evaluation of a treatment program for offenders.

Process evaluation is more important when more complex programs are evaluated. Many social programs comprise multiple elements and are delivered over an extended period, often by different providers in different areas. Because of this complexity, it is quite possible that the program as delivered is neither the same for all program recipients nor consistent with the formal program design.

The evaluation of D.A.R.E. by Research Triangle Institute researchers Christopher Ringwalt and others (1994) included a process evaluation designed to address these issues:

- Assess the organizational structure and operation of representative DARE programs nationwide.
- Review and assess the factors that contribute to the effective implementation of D.A.R.E. programs nationwide.
- Assess how D.A.R.E. and other school-based drug prevention programs are tailored to meet the needs of specific populations.

The process evaluation (they called it an "implementation assessment") was an ambitious research project with site visits and informal interviews, discussions, and surveys of D.A.R.E. program coordinators and advisers. These data indicated that D.A.R.E. was operating as designed and was running relatively smoothly. Drug prevention coordinators in D.A.R.E. school districts rated the program more highly than coordinators in districts with other alcohol and drug prevention programs rated theirs.

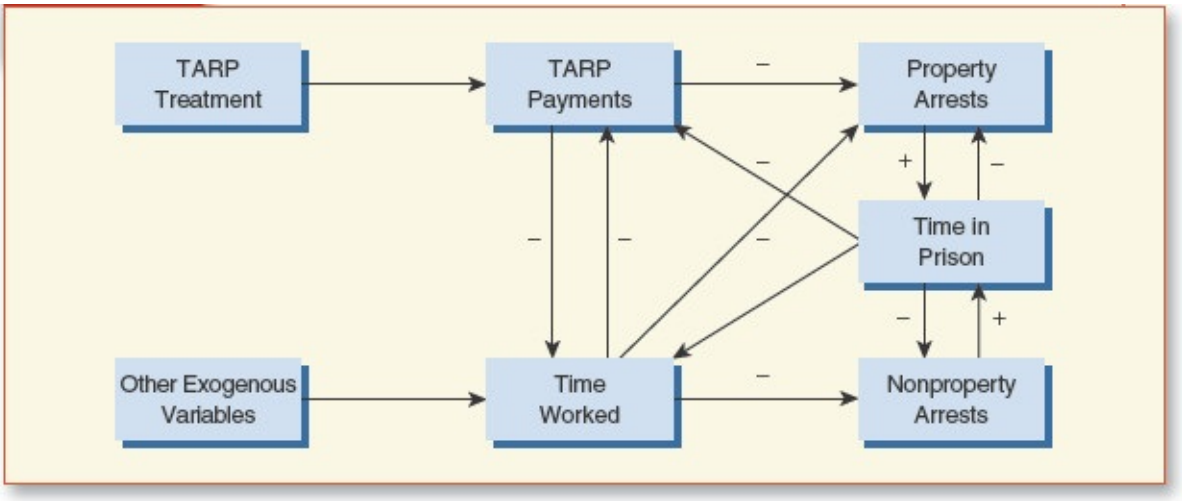
Process evaluation can also identify which specific part of the service delivery has the greatest impact. This, in turn, helps explain why the program has an effect and which conditions are required for the effect. (In Chapter 6, we described this as identifying the causal "mechanism.") In the D.A.R.E. research, site visits revealed an insufficient number of officers and a lack of Spanish-language D.A.R.E. books in a largely Hispanic school. At the same time, classroom observations indicated engaging presentations and active student participation (Ringwalt et al. 1994: 69, 70).

Process analysis of this sort can also help show how apparently clear findings may be incorrect. The apparently disappointing results of the Transitional Aid Research Project (TARP) provide an instructive lesson. TARP was a social experiment designed to determine whether financial aid during the transition from prison to the community would help released prisoners find employment and avoid returning to crime. Two thousand participants in Georgia and Texas were randomized to receive either a particular level of benefits over a particular period or no benefits (the control group). Initially, it seemed that the payments had no effect: The TARP treatment condition did not alter the rate of subsequent arrests for property or nonproperty crimes.

But this wasn't all there was to it. Peter Rossi tested a more elaborate causal model of TARP's effects, which is summarized in Exhibit 12.6. Participants who received TARP payments had more income to begin with and so had more to lose if they were arrested; therefore, they were less likely to commit crimes.

However, TARP payments also created a disincentive to work and, therefore, increased the time available in which to commit crimes. Thus, the positive direct effect of TARP (more to lose) was cancelled out by its negative indirect effect (more free time).

Exhibit 12.6 Model of TARP Effects



Source: Drake, Robert E., Gregory J. McHugo, Deborah R. Becker, William A. Anthony, and Robin E. Clark. 1996. The New Hampshire study of supported employment for people with severe mental illness. Journal of Consulting and Clinical Psychology 64:391–399. Used with permission.

Formative evaluation occurs when the evaluation findings are used to help shape and refine the program (Rossi & Freeman 1989), for instance by being incorporated into the initial development of the service program. Evaluation may then lead to changes in recruitment procedures, program delivery, or measurement tools (Patton 2002: 220).

You can see the *formative* element in the following government report on the performance of the Health Care Finance Administration (HCFA):

> While HCFA's performance report and plan indicate that it is making some progress toward achieving its Medicare program integrity outcome, progress is difficult to measure because of continual goal changes that are sometimes hard to track or that are made with insufficient explanation. Of

> the five fiscal year 2000 program integrity goals it discussed, HCFA reported that three were met, a fourth unmet goal was revised to reflect a new focus, and performance data for the fifth will not be available until mid-2001. HCFA plans to discontinue three of these goals. Although the federal share of Medicaid is projected to be $124 billion in fiscal year 2001, HCFA had no program integrity goal for Medicaid for fiscal year 2000. HCFA has since added a developmental goal concerning Medicaid payment accuracy. (U.S. Government Accounting Office 2001: 7)

Process evaluation can employ a wide range of indicators. Program coverage can be monitored through program records, participant surveys, community surveys, and analysis of users versus dropouts and ineligibles. Service delivery may be monitored through service records program staff complete, a management information system program administrators maintain, and program recipients' reports (Rossi & Freeman 1989).

Research|Social Impact Link

Read more about how evaluations are used.

Qualitative methods are often a key component of process evaluation studies because they can be used to elucidate and understand internal program dynamics —even those that were not anticipated (Patton 2002: 159; Posavac & Carey 1997). Qualitative researchers may develop detailed descriptions of how program participants engage with each other, how the program experience varies for different people, and how the program changes and evolves over time.

Process evaluation: Evaluation research that investigates the process of service delivery.

Research That Matters

Evaluation research on the Drug Abuse Resistance Education program (D.A.R.E.) in schools has long raised questions about its impact on drug abuse. However, program participation may positively affect students' attitudes toward the police. Amie Schuck at the University of Illinois at Chicago analyzed evaluation data already collected in a large randomized experiment that had tested the impact of D.A.R.E. in 12 pairs of urban and suburban schools in Illinois. Students' attitudes toward police had been measured with their answers to five questions asked in seven waves of data collection during a 7-year period.

Professor Schuck found that student attitudes toward the police became considerably more negative from the 5th and 6th grades, when the study began, to the 11th and 12th grades, when the study concluded. Other studies of youth attitudes toward the police have had similar results. However, participation in the D.A.R.E. program delayed the decline in attitudes toward the police, and then was associated with improved attitudes toward the police. This association was particularly strong for African American youth.

Source: Adapted from Schuck, Amie M. 2013. A life-course perspective on adolescents' attitudes to police: DARE, delinquency, and residential segregation. Journal of Research in Crime and Delinquency 50(4): 579–607.

Formative evaluation: Process evaluation that is used to shape and refine program operations.

Impact Analysis

The core questions of evaluation research are these: Did the program work? Did it have the intended result? This kind of research is variously called **impact analysis**, **impact evaluation**, or **summative evaluation**. Formally speaking, impact analysis compares what happened after a program was implemented with what *would* have happened had there been no program at all.

Think of the program—such as a new strategy for combating domestic violence or an income supplement—as an independent variable and the result it seeks as a dependent variable. The D.A.R.E. program (independent variable), for instance, tries to reduce drug use (dependent variable). If the program is present, we should expect less drug use. In a more elaborate study, we might have multiple values of the independent variable, for instance, comparing conditions of "no program," "D.A.R.E. program," and "other drug/alcohol education."

As in other areas of research, an experimental design is the preferred method for maximizing internal validity—that is, for making sure your causal claims about program impact are justified. Cases are assigned randomly to one or more experimental treatment groups and to a control group so that there is no systematic difference between the groups at the outset (see Chapter 6). The goal is to achieve a fair, unbiased test of the program itself so that differences between the types of people who are in the different groups do not influence judgment about the program's impact. It can be a difficult goal to achieve, however, because the usual practice in social programs is to let people decide for themselves whether they want to enter a program and to establish eligibility criteria that ensure that people who enter the program are different from those who do not (Boruch 1997). In either case, a selection bias is introduced.

But sometimes researchers are able to conduct well-controlled experiments. Robert Drake et al. (1996) evaluated the impact of two different approaches to providing employment services for people diagnosed with severe mental disorders, using a randomized experimental design. One approach, group skills training (GST), emphasized preemployment skills training and used separate agencies to provide vocational and mental health services. The other approach, individual placement and support (IPS), provided vocational and mental health services in a single program and placed people directly into jobs without

preemployment skills training. The researchers hypothesized that GST participants would be more likely to obtain jobs during the 18-month study period than would IPS participants.

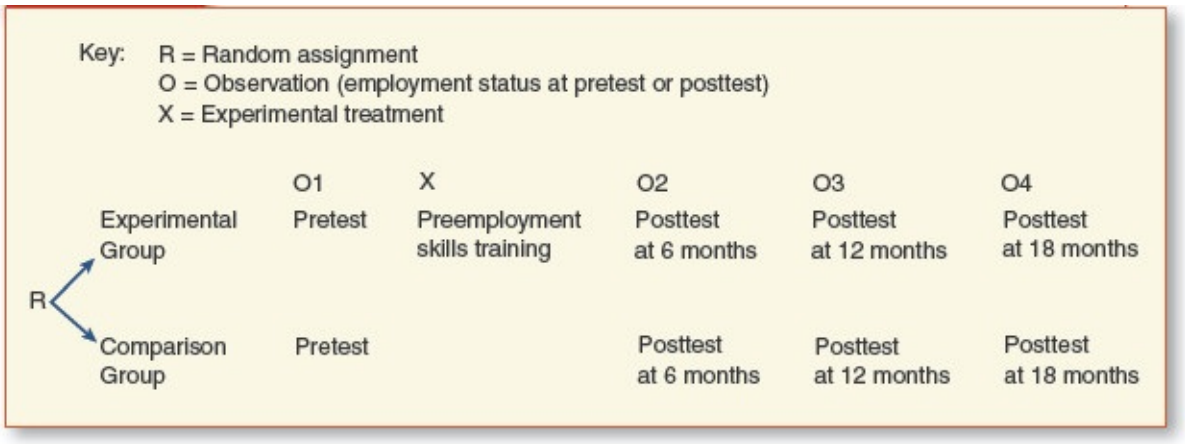
Their experimental design is depicted in Exhibit 12.7. Cases were assigned randomly to the two groups, and then

1. Both groups received a pretest.
2. One group received the experimental intervention (GST), and the other received the IPS approach.
3. Both groups received three posttests at 6, 12, and 18 months.

Contrary to the researchers' hypothesis, the IPS participants were twice as likely to obtain a competitive job as the GST participants were. The IPS participants also worked more hours and earned more total wages. Although this was not the outcome Drake et al. had anticipated, it was valuable information for policy makers and program planners—and the study was rigorously experimental.

Program impact also can be evaluated with quasi-experimental designs (see Chapter 6), nonexperimental designs, or field research methods without a randomized experimental design. If program participants can be compared with nonparticipants who are reasonably comparable except for their program participation, causal conclusions about program impact can still be made. However, researchers must evaluate carefully the likelihood that factors other than program participation might have resulted in the appearance of a program effect. For example, when a study at New York's maximum-security prison for women found that "education [i.e., classes] is found to lower risk of new arrest," the conclusions were immediately suspect: The research design did not ensure that the women who enrolled in the prison classes were the same as those who were not, "leaving open the possibility that the results were due, at least in part, to self-selection, with the women most motivated to avoid reincarceration being the ones who took the college classes" (Lewin 2001b: A18). Such nonequivalent control groups are often our only option, but you should be alert to their weaknesses.

Exhibit 12.7 Randomized Comparative Change Design: Employment Services for People With Severe Mental Disorders



Source: Orr, Larry L. 1999. Social experiments: Evaluating public programs with experimental methods. Thousand Oaks, CA: Sage, 224, Table 6.5. Reprinted with permission from SAGE Publications, Inc.

Impact analysis is an important undertaking that fully deserves the attention it has been given in government program funding requirements. However, you should realize that more rigorous evaluation designs are less likely to conclude that a program has the desired effect; as the standard of proof goes up, success is harder to demonstrate. The prevalence of "null findings" (or "we can't be sure it works") has led to a bit of gallows humor among evaluation researchers:

The Output/Outcome/Downstream Impact Blues

Donors often say,

And this is a fact,

Get out there and show us

Your impact

You must change peoples' lives

And help us take the credit

Or next time you want funding

You just might not get it.

So donors wake up

From your impossible dream.

You drop in your funding

A long way upstream.

The waters they flow,

They mingle, they blend

So how can you take credit

For what comes out in the end?

—Terry Smutylo, Director, Evaluation

International Development Research Centre

Ottawa, Canada

Source: Patton, Michael Quinn. 2002. *Qualitative research & evaluation methods*, 3rd ed. Thousand Oaks, CA: Sage, 154. Reprinted with permission from Terry Smutylo.

Impact analysis (impact evaluation or summative evaluation): Evaluation research that answers these questions: Did the program work? Did it have the intended result?

Efficiency Analysis

Finally, a program may be evaluated for how efficiently it provides its benefit; typically, financial measures are used. Are the program's financial benefits sufficient to offset the program's costs? The answer is provided by a **cost–benefit analysis**. How much does it cost to achieve a given effect? This answer is provided by a **cost-effectiveness analysis**. Program funders often require one or both of these types of **efficiency analysis**.

A cost–benefit analysis must (obviously) identify the specific costs and benefits to be studied, but my "benefit" may easily be your "cost." Program clients, for instance, will certainly have a different perspective on these issues than do taxpayers or program staff. Exhibit 12.8 lists factors that can be considered costs or benefits in a supported employment program from the standpoint of participants and taxpayers (Schalock & Butterworth 2000). Note that some anticipated impacts of the program (e.g., taxes and subsidies) are a cost to one group but a benefit to the other, and some impacts are not relevant to either.

Exhibit 12.8 Potential Costs and Benefits of a Social Program, by Beneficiary

Costs/Benefits	Perspective of Program Participants	Perspective of Rest of Society	Perspective of Entire Society*
Costs			
Operational costs of the program	0	–	–
Forgone leisure and home production	–	0	–
Benefits			
Earnings gains	+	0	+
Reduced costs of nonexperimental services	0	+	+
Transfers			
Reduced welfare benefits	–	+	0
Wage subsidies	+	-	0
Net benefits	±	±	±

Note: – = program costs; + = program benefits; ± = program costs and benefits; 0 = no program costs or benefits.

*Entire society = program participants + rest of society.

Video Link

Watch how cost-benefit analyses are important in making public policy decisions.

After potential costs and benefits have been identified, they must be measured. This need is highlighted in recent government programs (Campbell 2002):

> The Governmental Accounting Standards Board's (GASB) mission is to establish and improve standards of accounting and financial reporting for state and local governments in the United States. In June 1999, the GASB issued a major revision to current reporting requirements ("Statement 34"), which aims to provide information so citizens and other users can understand the financial position and cost of programs. (p. 1)

In addition to measuring services and their associated costs, a cost–benefit analysis must be able to make some type of estimation of how clients benefited from the program and what the economic value of this benefit was. A recent study of therapeutic communities provides a clear illustration. A *therapeutic community* (TC) is a method for treating substance abuse in which abusers participate in an intensive, structured living experience with other addicts who are attempting to stay sober. Because the treatment involves residential support as well as other types of services, it can be quite costly. Are those costs worth it?

Stanley Sacks and colleagues (2002) conducted a cost–benefit analysis of a modified TC in which 342 homeless, mentally ill chemical abusers were randomly assigned to either a TC or a "treatment-as-usual" comparison group. Employment status, criminal activity, and utilization of health care services were each measured for the 3 months before entering treatment and the 3 months after

treatment. Earnings from employment in each period were adjusted for costs incurred by criminal activity and utilization of health care services.

Was it worth it? The average cost of TC treatment for a client was $20,361. In comparison, the economic benefit (based on earnings) to the average TC client was $305,273, which declined to $273,698 after comparing post- to preprogram earnings. After adjusting for the cost of the program, the benefit was still $253,337. The resulting benefit–cost ratio was 13:1, although this ratio declined to only 5.2:1 after further adjustments (for cases with extreme values). Nonetheless, the TC program studied seems to have had a substantial benefit relative to its costs.

Cost–benefit analysis: A type of evaluation research that compares program costs with the economic value of program benefits.

Cost-effectiveness analysis: A type of evaluation research that compares program costs with actual program outcomes.

Efficiency analysis: A type of evaluation research that compares program costs with program effects. It can be either a cost–benefit analysis or a cost-effectiveness analysis.

Ethical Issues in Evaluation Research

Whenever you evaluate the needs of clients or analyze the impact of a program, you directly affect people's lives. Social workers want to believe their efforts matter; drug educators think they're preventing drug abuse. Homeless people have problems and may really appreciate the services an agency provides. Program administrators have bosses to please; foundations need big programs to fund; and domestic violence, for instance, is a real problem—and finding solutions to it matters. Participants and clients in social programs, then, are not just subjects eager to take part in your research; they care about your findings, deeply. This produces serious ethical as well as political challenges for the evaluation researcher (Boruch 1997: 13; Dentler 2002: 166).

There are many specific ethical challenges in evaluation research:

- How can confidentiality be preserved when the data are owned by a government agency or are subject to discovery in a legal proceeding?
- Who decides what burden an evaluation project can impose upon participants?
- Can a research decision legitimately be shaped by political considerations?
- Must findings be shared with all stakeholders or only with policy makers?
- Will a randomized experiment yield more defensible evidence than the alternatives?
- Will the results actually be used?

Is it fair to assign persons randomly to receive some social program or benefit? What fairer way is there to distribute scarce benefits than through a lottery? The State of Oregon has recently begun doing exactly this with some health care benefits (Yardley 2008). This is exactly the process that is involved in a randomized experimental design.

Interactive Exercises Link

Evaluation Research

The Health Research Extension Act of 1985 (Public Law 99–158) mandated that the Department of Health and Human Services require all research organizations

receiving federal funds to have an institutional review board (IRB) to assess all research for adherence to ethical practice guidelines. There are six federally mandated criteria (Boruch 1997):

- Are risks minimized?
- Are risks reasonable in relation to benefits?
- Is the selection of individuals equitable? (Randomization implies this.)
- Is informed consent given?
- Are the data monitored?
- Are privacy and confidentiality ensured? (pp. 29–33)

Evaluation researchers must consider these criteria before they even design a study. Subject confidentiality is particularly thorny because researchers, in general, are not usually exempted from providing evidence sought in legal proceedings. However, several federal statutes have been passed specifically to protect research data about certain vulnerable populations from legal disclosure requirements. For example, the Crime Control and Safe Streets Act (28CFR Part 11) includes the following stipulation (Boruch 1997):

> Copies of [research] information [about persons receiving services under the act or the subject of inquiries into criminal behavior] shall be immune from legal process and shall not, without the consent of the persons furnishing such information, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceedings. (p. 60)

When ethical standards can’t be met, modifications may be made in the study design. Several steps can be taken (Boruch 1997):

- Alter the group allocation ratios to minimize the number in the untreated control group.
- Use the minimum sample size required to be able to test the results adequately.
- Test just parts of new programs rather than entire programs.
- Compare treatments that vary in intensity (rather than presence or absence).
- Vary treatments between settings rather than among individuals within a setting. (pp. 67–68)

Conclusion

In social policy circles, hopes for evaluation research are high: Society would benefit from the programs that work well, that accomplish their goals, and that serve people who genuinely need them. At least that is the hope. Unfortunately, there are many obstacles to realizing this hope. Because social programs and the people who use them are complex, evaluation research designs can easily miss important outcomes or aspects of the program process. Because the many program stakeholders all have an interest in particular results from the evaluation, researchers can be subjected to an unusual level of cross-pressures and demands. Because the need to include program stakeholders in research decisions may undermine adherence to scientific standards, research designs can be weakened. Because program administrators may want to believe their programs really work well, researchers may be pressured to avoid null findings or, if they are not responsive, find their research reports ignored. Because the primary audience for evaluation research reports is program administrators, politicians, or members of the public, evaluation findings may need to be overly simplified, distorting the findings (Posavac & Carey 1997). Plenty of well-done evaluation research studies wind up in a recycling bin or hidden away in a file cabinet.

Encyclopedia Link

Read about the uses of evaluation research and its applications.

The rewards of evaluation research are often worth the risks, however. Evaluation research can provide social scientists with rare opportunities to study complex social processes, with real consequences, and to contribute to the public good. Although they may face unusual constraints on their research designs, most evaluation projects can also result in high-quality analyses and publications in reputable social science journals. In many respects, evaluation research is an idea whose time has come. We may never achieve Donald Campbell’s vision of an “experimenting society” (Campbell & Russo 1999) in which research is consistently used to evaluate new programs and to suggest constructive changes, but we are close enough to continue trying.

Key Terms

Cost–benefit analysis 307
Cost-effectiveness analysis 307
Efficiency analysis 307
Evaluability assessment 300
Feedback 292
Formative evaluation 305
Impact analysis (impact evaluation or summative evaluation) 305
Inputs 292
Integrative approaches (to evaluation) 296
Needs assessment 299
Outcomes 292
Outputs 292
Process evaluation 301
Program process 292
Program theory 294
Social science approaches (to evaluation) 295
Stakeholder approaches (to evaluation) 295
Stakeholders 293
Theory-driven evaluation 294

Highlights

- Evaluation research is social research that is conducted for a distinctive purpose: to investigate social programs.
- The development of evaluation research as a major enterprise followed on the heels of the expansion of the federal government during the Great Depression and World War II.
- The evaluation process can be modeled as a feedback system, with inputs entering the program, which generate outputs and then outcomes, which feed back to program stakeholders and affect program inputs.
- The process by which a program has an effect on outcomes is often treated as a "black box," but there is good reason to open the black box and investigate the process by which the program operates and produces, or fails to produce, an effect.
- A program theory may be developed before or after an investigation of the program process is completed. The theory can be either descriptive or prescriptive.
- The evaluation process as a whole, and the feedback process in particular, can only be understood in relation to the interests and perspectives of program stakeholders.
- Qualitative methods are useful in describing the process of program delivery.
- Multiple outcomes are often necessary to understand program effects.
- Evaluation research is research for a client, and its results may directly affect the services, treatments, or punishments that program users receive.
- There are five primary types of program evaluation: needs assessment, evaluability assessment, process evaluation (including formative evaluation), impact analysis (also termed *summative evaluation*), and efficiency (cost–benefit) analysis.
- Evaluation research raises complex ethical issues because it may involve withholding desired social benefits.

Student Study Site

Related Titles in Research Methods and Statistics From SAGE

Investigating the Social World: The Process and Practice of Research, Eighth Edition, by Russell K. Schutt

The Practice of Research in Social Work, Third Edition, by Rafael J. Engel and Russell K. Schutt

The Practice of Research in Criminology and Criminal Justice, Fifth Edition, by Ronet K. Bachman and Russell K. Schutt

Fundamentals of Research in Criminology and Criminal Justice, Third Edition, by Ronet K. Bachman and Russell K. Schutt

Designing Surveys: A Guide to Decisions and Procedures, Second Edition, by Ronald F. Czaja and Johnny Blair

A Guide to Qualitative Field Research, Second Edition, by Carol A. Bailey

Adventures in Social Research, Seventh Edition, by Earl Babbie, Fred Halley, William E. Wagner, III, and Jeanne Zaino

Adventures in Criminal Justice Research, Fourth Edition, by Kim Logio, George Dowdall, Earl Babbie, and Fred Halley

Social Statistics for a Diverse Society, Sixth Edition, by Chava Frankfort-Nachmias and Anna Leon-Guerrero

Essentials of Social Statistics for a Diverse Society by Anna Leon-Guerrero and Chava Frankfort-Nachmias

Research Methods in Psychology, Second Edition, by Paul G. Nestor and Russell K. Schutt

Research Methods in Education by Joseph Check and Russell K. Schutt

Fundamentals of Social Work Research, Second Edition, by Rafael J. Engel and Russell K. Schutt

Constructing Social Research by Charles C. Ragin and Lisa M. Amoroso

Using IBM® SPSS® Statistics for Social Statistics and Research Methods, Third Edition, by William E. Wagner, III

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore | Boston

The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. Would you prefer that evaluation researchers use a stakeholder or a social science approach? Compare and contrast these perspectives, and list at least four arguments for the one you favor.
2. Think of your primary health care provider as providing a “program” that should be evaluated. (If that makes you squeamish, you can focus on your college as the “program” instead.)

a. How would you describe the contents of the “black box” of program operations?

b. What program theory would specify how the program operates?

c. What would be the advantages and disadvantages of using qualitative methods to evaluate this program?

d. What would be the advantages and disadvantages of using quantitative methods?

e. Which approach would you prefer and why?

Finding Research

1. Inspect the website maintained by the Governmental Accounting Standards Board (www.seagov.org). Read and report on performance measurement in government as described in one of the case studies.
2. Describe the resources available for evaluation researchers at one of the following three websites: www.wmich.edu/evalctr/, http://www.innonet.org/, or www.worldbank.org/oed/.

Critiquing Research

1. Read and summarize an evaluation research report published in the *Evaluation and Program Planning* journal. Be sure to identify the type of evaluation research that is described.
2. Select one of the evaluation research studies described in this chapter, read the original report (book or article) about it, and review its adherence to the ethical guidelines for evaluation research. Which guidelines do you feel are most important? Which are most difficult to adhere to?

Doing Research

1. Propose a randomized experimental evaluation of a social program with which you are familiar. Include in your proposal a description of the program and its intended outcomes. Discuss the strengths and weaknesses of your proposed design.
2. Identify the key stakeholders in a local social or educational program. Interview several stakeholders to determine their goals for the program and what tools they use to assess goal achievement. Compare and contrast the views of each stakeholder, and try to account for any differences you find.

Ethics Questions

1. In the study of the housing alternatives by Schutt (2011), an ethnographer learned that a house resident was talking seriously about cutting himself. If you were the ethnographer, would you have immediately informed house staff about this? Would you have told anyone? What if the resident asked you not to tell anyone? In what circumstances would you feel it is ethical to take action to prevent the likelihood of a subject's harming himself or herself or others?
2. Is it ethical to assign people to receive some social benefit on a random basis? Form two teams and debate the ethics of the TARP randomized evaluation of welfare payments described in this chapter.

Video Interview Questions

1. Listen to the researcher interviews for Chapter 12 at edge.sagepub.com/chamblissmssw5e.
2. Why was this specific research study challenging?
3. How did the researchers come up with the “counterfactual” component of the study?

Chapter 13 Reviewing, Proposing, and Reporting Research

©iStockphoto.com/tumpikuja

Learning Objectives

1. Identify the strengths and weaknesses of alternative research designs.
2. Understand how to systematically evaluate research reports.
3. Explain the goals and challenges to keep in mind when writing a proposal.
4. Compare and contrast the different types of reports and know which to use to address specific needs.
5. Identify unique problems that must be overcome in writing student papers, theses, applied research reports, and journal articles.
6. List the major sections of a research report.
7. Understand the importance of revising and peer review in writing.
8. Identify major steps in the review of research reports.
9. Be aware of the problem of plagiarism

In a sense, we end this book where we began. As you begin writing up your findings, you can see the gaps in the research. While reviewing the literature—and finding where your own work fits in—you may discover more interesting possibilities or more exciting studies to be started. In the process of concluding each study, we almost naturally begin the next.

The primary goals of this chapter are to guide you in evaluating the research of other scholars, developing research proposals, and writing worthwhile reports of your own. We first discuss how to evaluate prior research—a necessary step before writing a research report or proposal. We then focus on writing research proposals and reports.

Comparing Research Designs

From different methods, we learn different things. Even when used to study the same social processes, the central features of experiments, surveys, qualitative methods, and evaluation research provide distinct perspectives. Comparing subjects randomly assigned to a treatment group and to a comparison group, asking standard questions of the members of a random sample, observing while participating in a natural social setting, or studying program impact involve markedly different decisions about measurement, causality, and generalizability. As you can see in Exhibit 13.1, not one of these methods can reasonably be graded as superior to the others in all respects, and each varies in its suitability to different research questions and goals. Choosing among them for a particular investigation requires consideration of the research problem, opportunities and resources, prior research, philosophical commitments, and research goals.

Exhibit 13.1 Comparison of Research Methods[a]

Design	Measurement Validity	Generalizability	Causal Validity
Experiments	+	–	+
Surveys	+	+	–/+[b]
Participant Observation	–/+[c]	–	–

a. A plus (+) sign indicates where a method is strong; a minus (–) sign indicates where a method is weak.

b. Surveys are a weaker design for identifying causal effects than true experiments, but use of statistical controls can strengthen causal arguments.

c. Reliability is low compared with surveys, and systematic evaluation of measurement validity is often not possible.

Encyclopedia Link

Read an overview of qualitative versus quantitative research designs.

Experimental designs are strongest for testing nomothetic causal hypotheses (lawlike explanations that identify a common influence on a number of cases or events). These designs are most appropriate for studies of treatment effects (see Chapter 6). Research questions that are believed to involve basic social psychological processes are most appealing for laboratory studies because the problem of generalizability is reduced. Random assignment reduces the possibility of preexisting differences between treatment and comparison groups to small, specifiable, chance levels, so many of the variables that might create a spurious association are controlled. Laboratory experiments permit unsurpassed control over conditions and are excellent for establishing internal validity (causality).

But experimental designs have weaknesses. For most laboratory experiments, people volunteer as subjects, but volunteers aren't like other people, so generalizability is not good. Ethical and practical constraints limit your treatments (for instance, you can't randomly assign race or social class). Although some processes may be the same for all people, so that generalizing from volunteer subjects will work, it's difficult to know in advance which processes are really invariant. Field experiments, although apparently more generalizable studies, allow for less control than lab experiments; hence, treatments may not be delivered as intended, or other influences may intrude (see Chapter 9). Also, field experiments typically require unusual access (e.g., permission to revise a school curriculum or change police department policy) and can be very expensive.

Surveys, because of their probability sampling and standardized questions, are excellent for generalizable descriptive studies of large populations (see Chapter 7). They can include a large number of variables, unlike experiments, so that potential spuriousness can be statistically controlled; therefore, surveys can be used readily to test hypothesized causal relationships. And because many closed-ended questions are available that have been used in previous studies, it's easy to find reliable measures of commonly used variables.

But surveys, too, have weaknesses. Survey questionnaires can measure only what respondents are willing to say; surveys might not uncover behavior or attitudes that are socially unacceptable. Survey questions, being standardized, may miss the nuances of a respondent's feelings or the complexities of an attitude; they lump together what may be interestingly different responses. Surveys rely on the truthfulness of respondents and on their accuracy in reporting (for instance, students are asked how many hours a week they study—Do they know? Is study time constant?).

Qualitative methods allow intensive measurement of new or developing concepts, subjective meanings, and causal mechanisms (see Chapter 9). In field research, a grounded theory approach helps you create and refine concepts and theories based on direct observation or in-depth interviewing. Interviewing reveals what people really mean by their ideas and allows you to explore their feelings at great length. How, exactly, social processes unfold over time can be explored using interviews and fieldwork. Qualitative methods can identify the multiple successive events that might have led to some outcome, thus identifying idiographic causal processes; qualitative methods are excellent for studying new or poorly understood settings and populations that seek to remain hidden. When exploratory questions are posed or new groups studied, qualitative methods are preferred.

But such intensive study is time consuming, so fewer cases can be examined. Single or a few cases or unique settings are interesting but don't produce generalizable results. Also, most researchers can't spend 6 months away from home doing a project. Open-ended interviews take time—not just the 1 or 2 hours of the interview itself but time in scheduling, in missed appointments, in travel to reach your subjects, and so on.

When qualitative methods can find real differences in an independent variable—for example, several different management styles in a manufacturing company—you can test nomothetic causal hypotheses. But the impossibility of controlling numerous possible extraneous influences makes qualitative methods a weak approach to hypothesis testing.

Reviewing Research

A good literature review is the foundation for a research proposal, both in identifying gaps in current knowledge and in considering how to design a research project. It is also important to review the literature before writing an article about the research findings—the latest findings on your topic should be checked, and prior research on new issues should be consulted. This section helps you learn how to review the research that you locate. First, we focus on the process of reviewing single articles; then, we explain how to combine reviews of single articles into an overall literature review.

Exhibit 13.2 lists the questions you should ask when critiquing a social research study, and the following paragraphs provide an example. This particular critique does not answer all of the review questions, nor does it provide complete answers to all these questions, but it gives you the basic idea. In any case, remember that your goal is to evaluate research projects as integrated wholes. In addition to considering how valid the measures were and whether the causal conclusions were justified, you must consider how the *measurement approach* might have affected the causal validity of the researcher's conclusions and how the *sampling strategy* might have altered the quality of measures. In other words, all the parts of a study affect each other. Our goal here is just to illustrate the process of critically thinking about a piece of research.

Case Study: “Night as Frontier”

A minor classic in sociological literature, Murray Melbin’s 1978 article “Night as Frontier” compares 20th-century extension of human activity into nighttime hours with 19th-century geographic expansion into the American West. Melbin argues that just as there was a “frontier lifestyle” in the Old West of cowboys, a similar style of behavior, particularly toward strangers, prevails among late-night inhabitants of contemporary U.S. cities. In developing this comparison of spatial frontiers with temporal frontiers, Melbin accomplished an insightful reconceptualization of how human beings live on a sparsely populated “frontier” of a different kind.

Suppose that you are a student of urban life and curious whether city dwellers, such as New Yorkers, are really as unfriendly and brusque as stereotypes portray them. Melbin’s article describes a number of field experiments, conducted entirely in Boston, to discover whether people were more or less helpful to others at nighttime than during the day. Perhaps you could use his findings. But was his research properly conducted?

Exhibit 13.2 Questions to Ask About a Research Article

In reading a research article, you want to know (a) What is the author's conclusion? and (b) Does the research presented adequately support that conclusion? The questions below will help you determine the answers.

I. Overall assessment of the article
   1. What is the basic question being posed?
   2. Is the theoretical approach appropriate?
   3. Is the literature review adequate?
   4. Does the research design suit the question?
   5. Is the study scientific in its fundamentals?
   6. Are the ethical issues adequately addressed?
   7. What are the key findings?

II. Detailed assessment
   1. What are the key concepts? Are they clearly defined?
   2. What are the main hypotheses?
   3. What are the main independent and dependent variables?
   4. Are the measurements valid?
   5. What are the units of analysis? Are they appropriate?
   6. Are any causal relationships successfully established?
   7. Is the effective sample (sampling plus response rate) representative?
   8. Does context matter to the causal relationship?

The Research Design

Melbin and his assistants conducted four different experiments, all designed to measure whether time of day affected people's willingness to be "helpful or friendly" to strangers. He drew partly on a sizable literature in this area conducted by social psychologists, but his studies were simpler in design than most psychology experiments. In most cases, he had one independent variable—time of day—and one dependent variable—how likely people were to be helpful or friendly. Melbin's assistants, using a detailed sampling procedure (sampling both times of the day and subjects), approached random people on streets in Boston (also sampled). In one study, the researchers asked for directions; in another, they requested that subjects answer several interview questions. In a third study, they observed customers' interactions with cashiers at grocery stores. Finally, they left keys, tagged with "Please return" and an address, in various locations. In each case, the independent variable was time of day (for instance, when subjects were approached or the key was dropped on the street); the dependent variable was whether people were cooperative (directions,

interviews), helpful (returning key), or friendly (smiling, conversational). A clear, simple coding scheme was used for all of these measures.

Encyclopedia Link

Read a review of research designs.

Analysis of the Design

Melbin's study was exploratory, designed to propose a new idea of how to understand nighttime in contemporary America. His experiments, therefore, were more in the manner of demonstrations—a first test of a new idea—than of continuing an established line of scientific research. Indeed, Melbin (1978) himself claimed to be advancing "the hypothesis that night is a frontier"; yet his experiments only test the idea that people at night are more "helpful and friendly" to strangers, which he argues is one of about a dozen characteristics of frontier communities.

But we can narrow our view to his specific question about helpfulness. His measures certainly have face validity, and in fact, in three of his four studies, people were indeed more friendly at night. And he didn't simply ask people if they would be helpful; he tested them in real situations in which they didn't know that it was an experiment. He also was open to surprises: In the "lost key" study, people were in fact *less* likely to return the key at night. Melbin realized that he had unintentionally slipped in another variable—whether the act of helpfulness was anonymous (the key study) or not (all the others). Only the community of face-to-face contact, he suggests, exists at night; help is not generally extended to those not part of the nighttime community. So the different trials also lend plausibility to his argument. He only studied city residents and only in Boston; it may be that the "nighttime community" exists only in urban settings, but an urban setting was a constant, not a variable, here.

There are at least two important problems in Melbin's design, despite its conscientious use of sampling, reliable coding procedures, and multiple measures. First, the studies don't really show that nighttime makes particular people more helpful and friendly; they show that people who are up at night—a self-selected group—are more helpful and friendly.

Perhaps the kind of people who prefer nightlife, not nighttime itself, is the true causal agent. And second, again, the studies were all conducted in a Northeastern city. Rural or suburban settings—a different context—could very well reveal different patterns.

An Overall Assessment

“Night as Frontier” certainly makes a persuasive argument with far more historical and theoretical detail than we’ve mentioned here. It tends to be research of the “exploratory” type, so its experiments are somewhat crude; neither the measures nor the studies themselves have been widely replicated. Ethically, the work is benign. Its main value may lie in the persuasiveness of the argument that nighttime is different than daytime and that the difference is much like the difference between densely settled areas and the old frontier West. For its conceptual insights, “Night as Frontier” deserves a respected place in the social science literature. In a detailed study of urban life and community, it may be helpful, but perhaps it is not fundamental.

Case Study: When Does Arrest Matter?

The goal of the literature review process is to integrate the results of your separate article reviews and develop an overall assessment of the implications of prior research. The integrated literature review should accomplish three goals (Hart 1998):

1. Summarize prior research.
2. Critique prior research.
3. Present pertinent conclusions. (pp. 186–187)

We'll discuss each of these goals in turn.

Summarize Prior Research

Your summary of prior research must focus on the particular research questions that you will address, but you may need also to provide some more general background. Carolyn Hoyle and Andrew Sanders (2000: 14) begin their *British Journal of Criminology* research article about mandatory arrest policies in domestic violence cases with what they term a "provocative" question: What is the point of making it a crime for men to assault their female partners and ex-partners? Hoyle and Sanders then review the different theories and supporting research that has justified different police policies: the "victim choice" position, the "pro-arrest" position, and the "victim empowerment" position. Finally, they review the research on the "controlling behaviors" of men that frames the specific research question on which they focus: how victims view the value of criminal justice interventions in their own cases (p. 15).

Audio Link

Listen to a clip about critiquing prior research.

Ask yourself three questions about your summary of the literature (Pyrczak 2005):

1. *Have you been selective?* If there have been more than a few prior

investigations of your research question, you will need to narrow your focus to the most relevant and highest-quality studies. Don't cite a large number of prior articles "just because they are there."
2. *Is the research up-to-date?* Be sure to include the latest research, not just the "classic" studies.
3. *Have you used direct quotes sparingly?* To focus your literature review, you need to express the key points from prior research in your own words. Use direct quotes only when they are essential for making an important point. (pp. 51–59)

Critique Prior Research

Evaluate the strengths and weaknesses of the prior research, answering the questions in Exhibit 13.2. You should select articles for review that reflect the work of credible authors in peer-reviewed journals who have been funded by reputable sources. Consider the following questions as you decide how much weight to give each article (Locke, Silverman, & Spirduso 1998):

1. *How was the report reviewed before its publication or release?* Articles published in academic journals go through a very rigorous review process, usually involving careful criticism and revision. Top "refereed" journals may accept only 10% of submitted articles, so they can be very selective. Dissertations go through a lengthy process of criticism and revision by a few members of the dissertation writer's home institution. A report released directly by a research organization is likely to have had only a limited review, although some research organizations maintain a rigorous internal review process. Papers presented at professional meetings may have had little prior review. Needless to say, more confidence can be placed in research results that have been subject to a more rigorous review.
2. *What is the author's reputation?* Reports by an author or team of authors who have published other work on the research question should be given somewhat greater credibility at the outset.
3. *Who funded and sponsored the research?* Major federal funding agencies and private foundations fund only research proposals that have been evaluated carefully and ranked highly by a panel of experts. These agencies also often monitor closely the progress of the research. This does not guarantee that every such project report is good, but it goes a long way toward ensuring some worthwhile products. However, research that is

funded by organizations that prefer a particular outcome should be given particularly close scrutiny. (pp. 37–44)

Present Pertinent Conclusions

Don’t leave the reader guessing about the implications of the prior research for your own investigation. Present the conclusions you draw from the research you have reviewed. As you do so, follow several simple guidelines (Pyrczak 2005):

Research|Social Impact Link

Read an article about why to be careful in interpreting and critiquing research.

- Distinguish clearly your own opinion of prior research from conclusions of the authors of the articles you have reviewed.
- Make it clear when your own approach is based on the theoretical framework you are using rather than on the results of prior research.
- Acknowledge the potential limitations of any empirical research project. Don’t emphasize problems in prior research that you can’t avoid either. (pp. 53–56)

Explain how the unanswered questions raised by prior research or the limitations of methods used in prior research make it important for you to conduct your own investigation (Fink 2005: 190–192).

A good example of how to conclude an integrated literature review is provided by an article based on the replication in Milwaukee of the Minneapolis Domestic Violence Experiment. For this article, Ray Paternoster and his colleagues (1997) sought to determine whether police officers’ use of fair procedures when arresting assault suspects would lessen the rate of subsequent domestic violence. Paternoster et al. concluded that there has been a major gap in the prior literature: “Even at the end of some seven experiments and millions of dollars, then, there is a great deal of ambiguity surrounding the question of how arrest impacts future spouse assault” (p. 164).

Specifically, the researchers noted that each of the seven experiments focused on the effect of arrest itself but ignored the possibility that “particular kinds of police *procedure* might inhibit the recurrence of spouse assault” (Paternoster et

al. 1997: 165).

So Paternoster and his colleagues (1997) grounded their new analysis in additional literature on procedural justice and concluded that their new analysis would be “the first study to examine the effect of fairness judgments regarding a punitive criminal sanction (arrest) on serious criminal behavior (assaulting one’s partner)” (p. 172).

Research That Matters

Cities across the United States have sought to reduce the toll of violent crimes by limiting access to guns. Strategies for controlling gun violence have ranged from gun buy-back programs, background checks, and safe storage laws to enhanced sentences for crimes committed with guns and community-based strategies. But do such strategies have the desired effect?

Matthew Makarios and Travis Pratt at the University of Cincinnati and Arizona State University, respectively, used meta-analysis to overcome these limitations. They were able to identify 27 research reports that included estimates of 172 effects of gun control programs. When they analyzed these studies together, they found that gun control programs tended to reduce violent crime, but only by a small amount. When they considered different types of gun control programs, they found that gun buy-back programs had no effect, whereas probation and community-oriented strategies had substantial effects—but the strongest effects occurred in studies with weaker research designs.

Source: Adapted from Makarios, Matthew D., and Travis C. Pratt. 2012. The effectiveness of policies and programs that attempt to reduce firearm violence: A meta-analysis. *Crime & Delinquency* 58(2): 222–244.

Proposing New Research

Be grateful for people who require you to write a formal research proposal—and even more for those who give you constructive feedback. Whether your proposal is written for a professor, a thesis committee, an organization seeking practical advice, or a government agency that funds basic research, the proposal will force you to set out a problem statement and a research plan. Too many research projects begin without a clear problem statement or with only the barest of notions about which variables must be measured or what the analysis should look like. Such projects often wander along, lurching from side to side, and then collapse entirely or just peter out with a report that is ignored—and should be. Even in circumstances when a proposal is not required, you should prepare one and present it to others for feedback. Just writing your ideas down will help you to see how they can be improved, and feedback in almost any form will help you to refine your plans.

A well-designed proposal can go a long way toward shaping the final research report and will make it easier to progress at later research stages (Locke, Spirduso, & Silverman 2000). Every research proposal should have at least six sections:

1. *An introductory statement of the research problem*, in which you clarify what it is that you are interested in studying
2. *A literature review*, in which you explain how your problem and plans build on what has already been reported in the literature on this topic
3. *A methodological plan*, detailing just how you will respond to the particular mix of opportunities and constraints you face
4. *A budget*, presenting a careful listing of the anticipated costs
5. *An ethics statement*, identifying human subjects issues in the research and how you will respond to them in an ethical fashion
6. *A statement of limitations*, reviewing weaknesses of the proposed research and presenting plans for minimizing their consequences

A research proposal also can be strengthened considerably by presenting a result of a pilot study of the research question. This might involve administering the proposed questionnaire to a small sample, conducting a preliminary version of the proposed experiment with a group of available subjects, or making

observations over a limited period in a setting like that proposed for a qualitative study. Careful presentation of the methods used in the pilot study and the problems that were encountered will impress anyone who reviews the proposal.

If your research proposal will be reviewed competitively, it must present a compelling rationale for funding. The research problem that you propose to study is crucial; its importance cannot be overstated (see Chapter 2). If you propose to test a hypothesis, be sure that it is one for which there are plausible alternatives, so your study isn't just a boring report of the obvious (Dawes 1995: 93).

Case Study: Community Health Workers and Cancer Clinical Trials

Particular academic departments, grant committees, and funding agencies will have specific proposal requirements. As an example, Exhibit 13.3 lists the primary required sections of the "Research Plan" for proposals to the National Institutes of Health (NIH), together with excerpts from a proposal by Russell Schutt, JudyAnn Bigby, and Lidia Schapira (2005) from two Harvard teaching hospitals submitted in this format to the National Cancer Institute (NCI) as part of a larger collaboration involving research and training at the University of Massachusetts Boston and the Dana Farber/Harvard Cancer Center (DF/HCC). The research plan (which is excerpted) must be preceded by a proposed budget, biographical sketches of project personnel, and a discussion of the available resources for the project. Appendixes may include research instruments, prior publications by the authors, and findings from related work.

Research|Social Impact Link

Read more about research proposals.

As you can see from the excerpts, the proposal was to study community health workers' (CHWs) knowledge of and orientations to cancer clinical trials and to then develop and test a training program for them about clinical trials. The proposal included two types of evaluation research: a needs assessment to learn about CHWs and clinical trials and an outcome assessment to identify changes in CHWs' knowledge and orientations as a result of participation in the training program. The NCI review committee (composed of experts in these issues) approved the project, and then after another administrative review, the project was awarded funds.

The reviewers recognized the proposal's strengths but also identified two issues that they believed had to be considered as the project was implemented. The issues were primarily methodological, related to validating the needs assessment tool and to using qualitative data.

Video Link

Watch a lecture on creating an effective research proposal.

> The primary goal of the training program is to help the CHWs effectively educate the communities they work with about the importance of clinical trials. An extensive program evaluation strategy has been included throughout the program development and implementation process. The evaluation will yield valuable information about CHWs' attitudes about clinical trials, how best to share this information with communities, about the effectiveness of community health workers to inform communities about clinical trials. This collaboration between DF/HCC and UMB represents a unique opportunity to build on the strengths of each institution to address a pressing problem that influences the persistence of cancer-related disparities.

Exhibit 13.3 A Grant Proposal to the National Cancer Institute

Community Health Workers and Cancer Clinical Trials

ABSTRACT

Disparities in cancer between subpopulations in the U.S. have been documented for several decades. One important area for intervention is the participation of underserved populations in cancer clinical trials. . . . Innovative community-based approaches are badly needed to affect these trends. This project will develop a clinical trials education training program for patient navigators and community health workers (CHWs). The primary goal of the training program is to help the CHWs effectively educate the communities they work with about the importance of clinical trials. An extensive program evaluation strategy has been included throughout the program development and implementation process. The evaluation will yield valuable information about CHWs' attitudes about clinical trials, how best to share this information with communities, about the effectiveness of community health workers to inform communities about clinical trials. . . .

RESEARCH PLAN

1. *Specific Aims*

    1. To develop a curriculum/program for training CHWs about clinical trials, so that they may educate the communities they work with about the importance of clinical trials
    2. To implement the training program with CHWs . . .
    3. To evaluate the impact of the training program . . .

2. *Background and Significance*

Risk, incidence, morbidity, and mortality for cancer in general and for some specific cancers are higher for blacks compared to whites, for poor persons compared to non-poor persons, and for rural residents compared to non-rural residents. Disparities have been documented across the cancer continuum ranging from risk factors and prevention to treatment and survival. The reasons for disparities in cancer treatment outcomes between different subpopulations are complex and many factors contribute. . . . One important area for intervention is the participation of underserved populations in cancer clinical trials.

Participation of minority populations in clinical trials is generally reported to be less frequent than participation of whites. . . .

Many barriers exist that prevent minority participation in clinical trials. . . . Most institutional committees charged with protecting human subjects do not adequately address all the concerns of these populations.

The federal government now requires that all persons involved in research with human subjects complete training on the principles of protection of human subjects. . . . many protections that have been instituted may actually serve as barriers to participation. For example, most IRBs now require extensive and highly detailed consent forms, which often use highly technical language and discuss procedures and concepts that are unfamiliar, overwhelming, and sometimes frightening.

Strategies to reverse the under-enrollment of minority and other underserved populations in clinical trials must address participant barriers, investigator barriers, and institutional barriers. We focus this proposal on an outreach strategy that will address some of the participant barriers . . .

An untapped resource in addressing the clinical trials accrual problem among underserved populations is the increasing number of CHWs [Community Health Workers] employed in many communities. . . . In the proposed project, we will develop a curriculum about clinical trials and train CHWs involved in several cancer screening and outreach programs to use or adapt the curriculum to educate several key communities about cancer clinical trials.

3. *Progress Report/Preliminary Studies*

C.1. Collaborators: This program is a collaboration between Dana Farber Harvard Cancer Center, specifically the Brigham and Women's Hospital (BWH) and Massachusetts General Hospital (MGH), and the University of Massachusetts, Boston (UMB). The study team includes Dr. JudyAnn Bigby from Brigham and Women's Hospital and Harvard Medical School, Dr. Lidia Schapira from Massachusetts General Hospital and Harvard Medical School, and Dr. Russell Schutt from the University of Massachusetts, Boston.

The proposed project will build on a program that was implemented at the Massachusetts General Hospital as part of an effort to address language and referral barriers for underserved populations. . . . Dr. Schapira and colleagues designed and implemented training programs for interpreters to increase their knowledge and skills.

Dr. Russell Schutt, at UMass Boston, will oversee the evaluation components of the project. Dr. Schutt is Professor of Sociology and Director, Graduate Program in Applied Sociology at UMass Boston and he is also Lecturer on Sociology in the Department of Psychiatry (MMHC/BID) at the Harvard Medical School. Dr. Schutt has extensive experience in

(Continued)

(Continued)

evaluation research and is the author of a leading research methods text in sociology (with versions adapted for social work, criminal justice, and undergraduate institutions). He has also designed ancillary training materials in research methods and has published more than 50 research articles and book chapters. He is co-investigator on the Women's Health Network Evaluation Project, an evaluation of the Mass. Department of Public Health case management program funded by the CDC's National Breast and Cervical Cancer Early Detection Program. He is also principal investigator . . . "Reviewing the Past, Planning the Future" at the Harvard Medical School. This project is recruiting a large team of health policy experts to review research about the Women's Health Network project and to ensure the most effective program operations. Dr. Schutt plays a key role in this program, as evaluation activities are incorporated throughout the curriculum and training development and implementation process, and are iterative in nature. We view ongoing evaluation as a critical component . . . .

4. *Research Design and Methods*

. . . During Year 1, the curriculum will be revised to meet the needs of a variety of CHWs, . . . Representatives from these community programs will participate in the development of the curriculum. We will pilot test the training program, and then revise it as needed (year 2). . . . UMB will evaluate the development of the training, the training itself, . . . and conclude with an outcome analysis of the program's impact. These evaluation activities will help to design the program curriculum and to implement the most effective program components.

D.2. Curriculum Development and Training

We propose to develop a curriculum designed specifically to meet the learning needs of the CHWs, and to provide state-of-the-art knowledge of the process and language of clinical trials. . . . Our efforts to develop an effective training program will involve four steps: 1) needs assessment; 2) curriculum development; 3) pilot testing of the training program; and 4) revision of the curriculum and training program. Each of these steps is described below.

D.2.1. Needs Assessment: . . . The first phase of the project will include a needs assessment in order to identify the level of understanding and knowledge of community workers with respect to clinical trials. . . . First, we will conduct two focus groups with CHWs to probe the attitudes and beliefs about clinical trials, their experiences with community outreach, and their impressions of client orientations . . . .Second, . . . ten in-depth interviews (approximately one hour in length) will be conducted with selected health workers .... These interviews will be designed to provide more details about issues raised in the focus groups . . . Third, a short structured survey will be designed to assess the backgrounds, attitudes and experiences of all CHWs involved in the project. This survey will include a measure of understanding of and attitudes toward clinical trials as well as information on the languages and cultural backgrounds of the CHWs. . . .

D.3.3. Program Evaluation: There will be several strategies utilized for evaluating the proposed program. First, an impact analysis will measure the change in CHW's understanding of and attitudes toward clinical trials. A structured survey . . .related to community education and clinical trials will be administered to participants prior to and following each training. . . . A measure of satisfaction with the training . . .

5. *Human Subjects*

E.1. Risks to Subjects: The risks of participation are minimal. The primary risk is the potential for loss of confidentiality.

E.2. Adequacy of Protection Against Risks: Confidentiality will be maintained by numerically coding data, . . . All information obtained from subjects will be accessible only to research staff.

E.3. Potential Benefits of the Proposed Research to Subjects: The proposed program evaluation will help . . . develop a community-based clinical trials education program . . . responsive to the needs . . . and reflects the language and values of the community.

E.4. Importance of the Knowledge to be Gained: . . . This project will help to address disparities in knowledge related to clinical trials, and . . . may impact on differential enrollment among minority cancer patients in clinical trials.

E.5. Women, Ethnic Minority, and Child Inclusion: All participants in the present investigation will be adults. We anticipate that the majority of participants will be women, . . .

E.5.1. Minority recruitment plan: We will work with all community health workers employed by specific programs .... The majority . . . are members of minority groups.

E.6. Risks Compared to Benefits: The benefits of the proposed study outweigh the potential risks. The knowledge gained will be substantial, and the risks are few and largely preventable. . .

E.7. Data Safety Monitoring: A data safety monitoring plan (DSMP) has been developed for this study. . . . All investigator-level staff members have completed the NIH human subject's certification as required. This is a minimal risk study, and thus we do not anticipate safety concerns.

Co-Leaders: Members of the investigative team have clearly delineated responsibilities based on their areas of expertise. . . .

Institutional Environment: The institutional environment at HMS is excellent and several collaborations currently exist that will facilitate

recruitment for this pilot project. . . .

Merit/Importance: The purpose of this pilot is to take advantage of the popular community health worker (CHW) model to develop, implement, and evaluate a curriculum/program for training CHWs to educate the communities in which they work about the importance of clinical trials. The rationale is that CHWs, with adequate training, could help community residents overcome certain barriers to clinical trials participation (e.g. lack of knowledge, mistrust, limited understanding, limited access to accurate/reliable information). This project builds on prior experiences training medical interpreters about clinical trials. The project will include 1) curriculum development (following a needs assessment via focus groups and in-depth interviews) that will include pilot testing and revisions, 2) implementation (training) and 3) program evaluation. The pilot is well described, with expected outcome and measurement strategies addressed. Examples of curricular content are provided. The evaluation plan will include both process and outcome measures. Plans to observe community education programs offered by the newly trained CHWs are also included. Potential challenges are acknowledged and incorporated into the training program (e.g., strategies to help CHWs maintain a focus on clinical trials education in their encounters and community education efforts). (Herberman 2005: 16–17)

Although the research plan is nicely laid out, there are a few remaining questions:

1. How will the survey designed to assess backgrounds, attitudes, and experience of CHWs be validated?
2. Will qualitative data from the CHWs be used to inform curricular development and, if so, in what ways?

. . . Future Potential: If successful, the curriculum could be implemented in other locations. The investigators also plan to evaluate the adaptability of the training to a train-the-trainer model. Given the popularity of the CHW model particularly in minority communities, this is a timely educational proposal.

NIH review committees reject most research proposals, require a revision before the others are recommended for funding, and do not actually fund many of even the meritorious proposals, so NCI's decision about this proposal was very welcome news. If you get the impression that researchers cannot afford to leave any stone unturned in working through procedures in an NIH proposal, you are right. It is very difficult to convince a government agency that a research project is worth spending money on. And that is as it should be: Your tax dollars should be used only for research that has a high likelihood of yielding findings that are valid and useful. But even when you are proposing a smaller project to a more generous funding source—or even presenting a proposal to your professor—you should scrutinize the proposal carefully before submission and ask others to comment on it. Other people will often think of issues you neglected to consider, and you should allow yourself time to think about these issues and to reread and redraft the proposal. Besides, you will get no credit for having thrown together a proposal as best you could in the face of an impossible submission deadline.

When you develop a research proposal, it will help to work through each of the issues in Exhibit 13.4 (also see Herek 1995). It is too easy to omit important details and to avoid being self-critical while rushing to put a proposal together. However, it is painful to have a proposal rejected (or to receive a low grade). Better to make sure the proposal covers what it should and confronts the tough issues that reviewers (or your professor) will be sure to spot.

The points in Exhibit 13.4 can serve as a map to preceding chapters in this book and as a checklist of decisions that must be made throughout any research project. The points are organized in five sections, each concluding with a *checkpoint* at which you should consider whether to proceed with the research as planned, modify the plans, or stop the project altogether. The sequential ordering of these questions obscures a bit the way in which they should be answered: not as single questions, one at a time, but as a unit—first as five separate stages and then as a whole. Feel free to change your answers to earlier questions on the basis of your answers to later questions.

A brief review of how the questions in Exhibit 13.4 might be answered with respect to the proposal to the National Cancer Institute by Schutt and colleagues (2005) should help you to review your own work. The research question concerned the need for and efficacy of a training program about cancer clinical trials, an evaluation research question (Question 1). This problem certainly was suitable for social research, and the funds we requested were judged to be

adequate ($66,204 for the evaluation component) (Question 2). Prior research demonstrated a need for the investigation and the potential for our training program. Schutt's own prior research (Estabrook, Schutt, & Woodford 2008; Schutt, Cruz, & Woodford 2008; Schutt, Fawcett et al. 2010) helped indicate the potential for the new proposed research (Question 3). The proposal did not make a direct connection to social theory—a common deficit in evaluation research proposals—but did emphasize relevant prior research (Question 4). The evaluation research plan had both inductive (needs assessment) and deductive (program impact) elements (Question 5). The review of research guidelines continued until submission, and Schutt and his colleagues felt that their proposal considered each (Question 6). So it seemed reasonable to continue to develop the proposal (Checkpoint 1).

Measures would be developed through coding of qualitative data collected in focus groups and intensive interviews, analysis of survey data, and observations of training sessions. The specific measures in the quantitative survey instruments and in the observational protocol had been used in prior research and some evidence had been presented suggesting their validity (Question 7). This pilot study was relatively weak in generalizability because Schutt and colleagues had to plan on studying an availability sample of community health workers (Question 8). Their needs assessment would involve only cross-sectional survey data, so they could only plan a strategy of multivariate statistical controls to test hypotheses about influences on knowledge and orientations. Their impact analysis was to include a before-and-after test to identify changes in individuals' knowledge and orientations, so their conclusion about an effect of the training program would have a somewhat stronger basis (Questions 9, 10, 11). They did not have a comparison for the impact analysis that was not exposed to the training they planned to develop, so endogenous change and external events were potential sources of causal invalidity. There was also a special basis for concern about an interaction of selection and treatment because those who agreed to participate in the training program could have been more open to change than were those who didn't participate; without randomized assignment to the training program or a comparison group, the researchers could not be sure (Question 12). Despite some weaknesses, the potential value of the training program they were to develop, and the possibility of more rigorous tests of its value in the future, encouraged Schutt and his colleagues to continue with their plan (Checkpoint 2).

The use of mixed-methods design was appropriate to the needs assessment

portion of their research. A randomized experimental design would have been preferable for the impact analysis, but it was not possible to plan such a study within the limitations of their budget and time (Questions 13, 14). Neither Schutt and coresearchers nor the reviewers identified ethical concerns in the project, other than preserving the confidentiality of data collected. The noninvasive nature of their methods and their focus on issues concerning community health workers' job-related concerns meant that there was little potential for harm resulting from participation in their research. Neither the University of Massachusetts Boston's Institutional Review Board (IRB) nor the Dana Farber/Harvard Cancer Center's IRB found there to be ethnical concerns about their plans (Question 15). Implementing the research plan seemed justified (Checkpoint 3).

Exhibit 13.4 Decisions in Research Design

PROBLEM FORMULATION (Chapters 1–2)

1. Developing a research question
2. Assessing researchability of the problem
3. Consulting prior research
4. Relating to social theory
5. Choosing an approach: Deductive? Inductive? Descriptive?
6. Reviewing research guidelines

CHECKPOINT 1

Alternatives:
- Continue as planned.
- Modify the plan.
- STOP. Abandon the plan.

RESEARCH VALIDITY (Chapters 4–6)

7. Establishing measurement validity
8. Establishing generalizability
9. Establishing causality
10. Data required: Longitudinal or cross-sectional?
11. Units of analysis: Individuals or groups?
12. What are major possible sources of causal invalidity?

CHECKPOINT 2

Alternatives:
- Continue as planned.
- Modify the plan.
- STOP. Abandon the plan.

RESEARCH DESIGN (Chapters 6–11)

13. Choosing a research design, such as survey or participant observation
14. Specifying the research plan: Types of experiments, surveys, observations, etc.
15. Assessing ethical concerns

CHECKPOINT 3

Alternatives:
- Continue as planned.
- Modify the plan.
- STOP. Abandon the plan.

DATA ANALYSIS (Chapter 8 and 10)

16. Choosing statistics, such as frequencies, cross-tabulation, etc.

CHECKPOINT 4

Alternatives:
- Continue as planned.
- Modify the plan.
- STOP. Abandon the plan.

REVIEWING, PROPOSING, AND REPORTING RESEARCH (Chapter 13)

17. Organizing the text
18. Reviewing ethical and practical constraints

CHECKPOINT 5

Alternatives:
- Continue as planned.
- Modify the plan.
- STOP. Abandon the plan.

Schutt and his colleagues expected to use descriptive univariate and multivariate statistics for the analysis of their needs assessment data, as well as a grounded theory approach for the analysis of their qualitative data. They planned to use inferential statistics to test for differences in mean knowledge and orientations before and after the training program (Question 16). They organized their proposal in the sections required by the National Institutes of Health. Before reporting their results, they first wrote a comprehensive research report on the needs assessment (Schutt et al. 2008), and they subsequently published separate

articles in peer-reviewed journals on the needs assessment (Schutt, Schapira et al. 2010) and on the impact analysis (Schapira & Schutt 2011) (Question 17). They continued to review ethical and practical constraints throughout the project, but they encountered few unexpected obstacles and were able to overcome the challenges they did confront in recruitment for the training (Question 18).

Careers and Research

Ruth Westby, MPH, Research Associate, IFC International

Source: Ruth Westby

For Ruth Westby, research—particularly public health research—means the chance to make new discoveries that affect people's lives by improving community health. She has studied how programs for disadvantaged and underserved groups are implemented and whether they have meaningful health impacts.

Westby was inspired to pursue a career in clinical research after her father died from cancer shortly after she received her BA from Emory University. After a few years of working with sick individuals on clinical trials, she decided to focus on public health so that she could look toward *preventing* disease. She sought out skill-based research courses and then internships that would help her use those skills as a graduate student. One such internship, at the Centers for Disease Control and Prevention, led to coauthored journal articles and a presentation at a large conference. In this way, Westby was exposed to opportunities that cemented her passion for public health research and provided a job in which every day at work is different and evokes a sense of pride.

Westby's research job also has kept her learning new research methods. She has already been exposed to systematic literature reviews, secondary data analyses, quantitative and qualitative data collection and analyses, and program evaluation. She finds program evaluation particularly rewarding, as she studies how programs are implemented and whether they have meaningful health impacts on disadvantaged populations.

If she could give current students advice, it would be to take advantage of mentors, faculty members, and anyone who is willing to help you learn:

> I've seen firsthand the advantages of getting to know faculty members on a personal level, networking and interning at institutions where I might want to work later, and using new research skills outside of class. Doing all of these things taught me so much more than if I had just attended lectures and read my textbooks. By the time I graduated from graduate school, I felt much more competent and set up for success than after college. In the long run, those relationships and experiences will mean just as much, if not more, than your GPA or course schedule.

Reporting Research

The goal of research is not just to discover something, but also to communicate that discovery to a larger audience: other social scientists, government officials, your teachers, the general public—perhaps several of these audiences. Whatever the study's particular outcome, if the research report enables the intended audience to comprehend the results and learn from them, the research can be judged a success. If the intended audience is not able to learn about the study's results, the research should be judged a failure—no matter how expensive the research, how sophisticated its design, or how much of yourself you invested in it.

You began writing your research report when you worked on the research proposal, and you will find that the final report is much easier to write, and more adequate, if you write more material for it as you work out issues during the project. It is very disappointing to discover that something important was left out when it is too late to do anything about it. And we don't need to point out that students (and professional researchers) often leave final papers (and reports) until the last possible minute (often for understandable reasons, including other coursework and job or family responsibilities). But be forewarned: *The last-minute approach does not work for research reports.*

Journal Link

Read an article whose conclusions are difficult to report because of sensitive issues and topics.

Writing and Organizing

A successful report must be well organized and clearly written. Getting to such a product is a difficult but not impossible goal. Consider the following principles formulated by experienced writers (Booth, Colomb, & Williams 1995):

- Respect the complexity of the task and don’t expect to write a polished draft in a linear fashion. Your thinking will develop as you write, causing you to reorganize and rewrite.
- Leave enough time for dead ends, restarts, revisions, and so on and accept the fact that you will discard much of what you write.
- Write as fast as you comfortably can. Don’t worry about spelling, grammar, and so on until you are polishing things up.
- Ask anyone you trust for reactions to what you have written.
- Write as you go along, so you have notes and report segments drafted even before you focus on writing the report. (pp. 150–151)

It is important to outline a report before writing it, but neither the organization of the report nor the first written draft should be considered fixed. As you write, you will get new ideas about how to organize the report. Try them out. As you review the first draft, you will see many ways to improve your writing. Focus particularly on how to shorten and clarify your statements. Make sure that each paragraph concerns only one topic. Remember the golden rule of good writing: Writing is revising!

You can ease the burden of writing in several ways:

Video Link

Watch advice on how to anticipate writing, organizing, and presenting the results of social research.

- Draw on the research proposal and on project notes. You aren’t starting from scratch; you have all the material you’ve written during the course of the project.
- Refine your word-processing skills on the computer so that you can use the most efficient techniques when reorganizing and editing.

- Seek criticism from friends, teachers, or other research consumers before you turn in the final product. They will alert you to problems in the research or the writing.

We often find it helpful to use **reverse outlining**. After you have written a first draft, read through the draft, noting down the key ideas as they come up. Do those notes reflect your original outline, or did you go astray? Are the paragraphs clean? How could your organization be improved?

Most important, leave yourself enough time so that you *can* revise, several times if possible, before turning in the final draft.

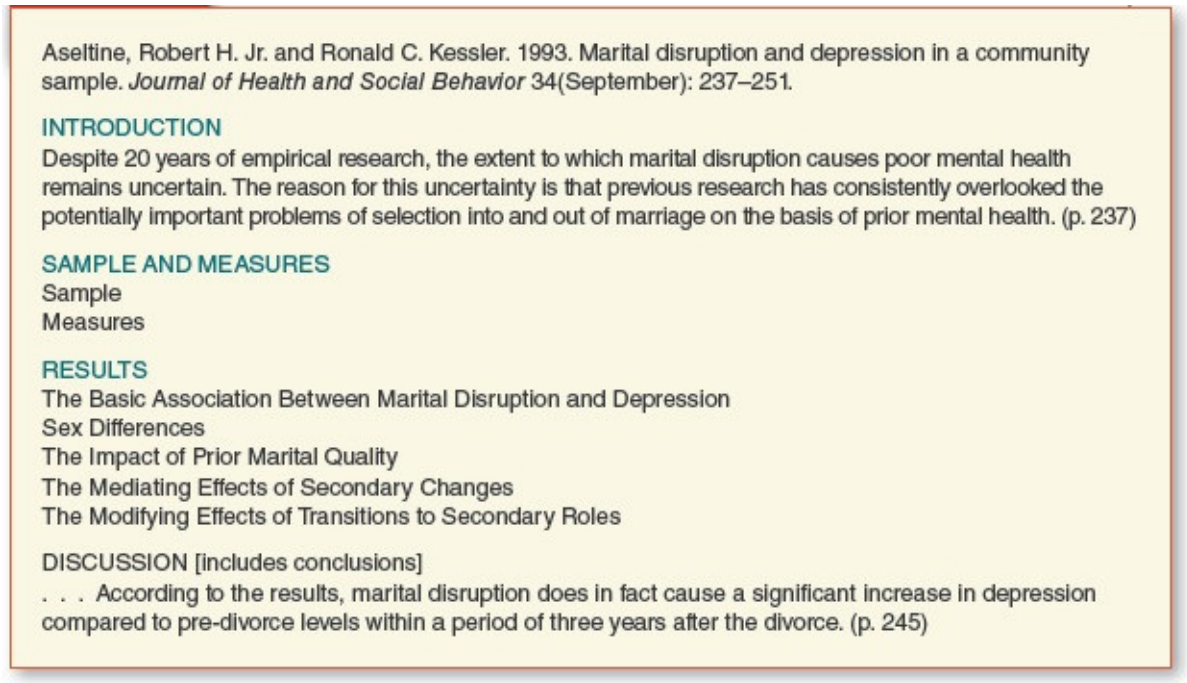
You can find more detailed reviews of writing techniques in Howard Becker (1986), Wayne Booth, Gregory Colomb, and Joseph Williams (1995), Carolyn Mullins (1977), William Strunk Jr. and E. B. White (1979), and Kate Turabian (1967).

Your report should be clearly organized into sections, probably following a standard format that readers will immediately understand. Any research report should include an *introductory statement of the research problem*, a *literature review*, and a *methodology section.* These same three sections should begin a research proposal. In addition, a research report must include a *findings section* with pertinent data displays. A *discussion section* may be used to interpret the findings and review the support for the study's hypotheses. A *conclusions section* should summarize the findings and draw implications for the theoretical framework used. Any weaknesses in the research design and ways to improve future research should be identified in this section. Compelling foci for additional research on the research question also should be noted. Most journals require a short abstract at the beginning that summarizes the research question and findings. A *bibliography* is also necessary. Depending on how the report is being published, *appendixes* containing the instruments used and specific information on the measures also may be included.

Exhibit 13.5 presents an outline of the sections in an academic journal article with some illustrative quotes. The article's introduction highlights the importance of the problem selected—the relation between marital disruption (divorce) and depression. The introduction also states clearly the gap in the research literature that the article is meant to fill—the untested possibility that depression might cause marital disruption rather than, or in addition to, marital

disruption causing depression. The findings section (labeled “Results”) begins by presenting the basic association between marital disruption and depression. Then the section elaborates on this association by examining sex differences, the impact of prior marital quality, and various mediating and modifying effects. As indicated in the combined discussion and conclusions section, the analysis shows that marital disruption does indeed increase depression and specifies the time frame (3 years) during which this effect occurs.

Exhibit 13.5 Sections in a Journal Article



Aseltine, Robert H. Jr. and Ronald C. Kessler. 1993. Marital disruption and depression in a community sample. *Journal of Health and Social Behavior* 34(September): 237–251.

INTRODUCTION

Despite 20 years of empirical research, the extent to which marital disruption causes poor mental health remains uncertain. The reason for this uncertainty is that previous research has consistently overlooked the potentially important problems of selection into and out of marriage on the basis of prior mental health. (p. 237)

SAMPLE AND MEASURES

Sample
Measures

RESULTS

The Basic Association Between Marital Disruption and Depression
Sex Differences
The Impact of Prior Marital Quality
The Mediating Effects of Secondary Changes
The Modifying Effects of Transitions to Secondary Roles

DISCUSSION [includes conclusions]

. . . According to the results, marital disruption does in fact cause a significant increase in depression compared to pre-divorce levels within a period of three years after the divorce. (p. 245)

Source: Aseltine Jr., Robert H., and Ronald C. Kessler. 1993. Marital disruption and depression in a community sample. Journal of Health and Social Behavior 34 (September): 237–251.

Audio Link

Listen to a podcast on understanding the elements of a journal article.

These basic report sections present research results well, but many research reports include subsections tailored to the issues and stages in the specific study

being reported. Lengthy applied reports on elaborate research projects may be organized around the research project's different stages or foci.

The material that can be termed the **front matter** and the **back matter** of an applied report also is important. Applied reports usually begin with an executive summary: a summary list of the study's main findings, often with bullet points. Appendixes, the back matter, may present tables containing supporting data that were not discussed in the body of the report. Applied research reports also often append a copy of the research instrument(s).

For instance, Exhibit 13.6 outlines the sections in an applied research report. This particular report was mandated by the California State Legislature to review a state-funded program for the homeless mentally disabled. The goals of the report are described as both description and evaluation. The body of the report presents findings on the number and characteristics of homeless persons and on the operations of the state-funded program in each of 17 counties. The discussion section highlights service needs that are not being met. Nine appendixes then provide details on the study methodology and the counties studied.

An important principle for the researcher writing for a nonacademic audience is to make the findings and conclusions engaging and clear. You can see how Schutt did this in a report from a class research project designed with his graduate methods students (and in collaboration with several faculty knowledgeable about substance abuse) (Exhibit 13.7). These report excerpts indicate how he summarized key findings in an executive summary (Schutt et al. 1996: iv), emphasized the importance of the research in the introduction (p. 1), used formatting and graphing to draw attention to particular findings in the body of the text (p. 5), and tailored recommendations to his own university context (p. 26).

A well-written research report requires (to be just a bit melodramatic) blood, sweat, and tears—and more time than you may at first anticipate. But writing one report will help you write the next report. And the issues you consider, if you approach your writing critically, will be sure to improve your subsequent research projects and sharpen your evaluations of other investigators' research projects.

Exhibit 13.6 Sections in an Applied Report

Vernez, Georges, M. Audrey Burnam, Elizabeth A. McGlynn, Sally Trude, and Brian S. Mittman. 1988. *Review of California's program for the homeless Mentally Disabled.* Santa Monica, CA: RAND.

SUMMARY

In 1986, the California State Legislature mandated an independent review of the HMD programs that the counties had established with the state funds. The review was to determine the accountability of funds; describe the demographic and mental disorder characteristics of persons served; and assess the effectiveness of the program. This report describes the results of that review. (p. v)

INTRODUCTION

Background
California's Mental Health Services Act of 1985 . . . allocated $20 million annually to the state's 58 counties to support a wide range of services, from basic needs to rehabilitation. (pp. 1–2)
Study Objectives
Organization of the Report

HMD PROGRAM DESCRIPTION AND STUDY METHODOLOGY

The HMD Program
Study Design and Methods
Study Limitations

COUNTING AND CHARACTERIZING THE HOMELESS

Estimating the Number of Homeless People
Characteristics of the Homeless Population

THE HMD PROGRAM IN 17 COUNTIES

Service Priorities
Delivery of Services
Implementation Progress
Selected Outcomes
Effects on the Community and on County Service Agencies
Service Gaps

DISCUSSION

Underserved Groups of HMD
Gaps in Continuity of Care
A particularly large gap in the continuum of care is the lack of specialized housing alternatives for the mentally disabled. The nature of chronic mental illness limits the ability of these individuals to live completely independently. But their housing needs may change, and board-and-care facilities that are acceptable during some periods of their lives may become unacceptable at other times. (p. 57)
Improved Service Delivery
Issues for Further Research

Appendixes

A. SELECTION OF 17 SAMPLED COUNTIES
B. QUESTIONNAIRE FOR SURVEY OF THE HOMELESS
C. GUIDELINES FOR CASE STUDIES
D. INTERVIEW INSTRUMENTS FOR TELEPHONE SURVEY
E. HOMELESS STUDY SAMPLING DESIGN, ENUMERATION, AND SURVEY WEIGHTS
F. HOMELESS SURVEY FIELD PROCEDURES
G. SHORT SCREENER FOR MENTAL AND SUBSTANCE USE DISORDERS
H. CHARACTERISTICS OF THE COUNTIES AND THEIR HMD-FUNDED PROGRAMS
I. CASE STUDIES FOR FOUR COUNTIES' HMD PROGRAMS

Source: Vernez, Georges M., Audrey Burnam, Elizabeth A. McGlynn, Sally Trude, and Brian S. Mirttman. 1988. Review of California's program for the homeless mentally disabled (R-3631- CDMH). Santa Monica, CA: RAND. Reprinted with permission.

Exhibit 13.7 Student Substance Abuse, Report Excerpts

EXECUTIVE SUMMARY

- Rates of substance abuse were somewhat lower at UMass–Boston than among nationally selected samples of college students.
- Two-thirds of the respondents reported at least one close family member whose drinking or drug use had ever been of concern to them—one-third reported a high level of concern.
- Most students perceived substantial risk of harm due to illicit drug use, but just one-quarter thought alcohol use posed a great risk of harm.

INTRODUCTION

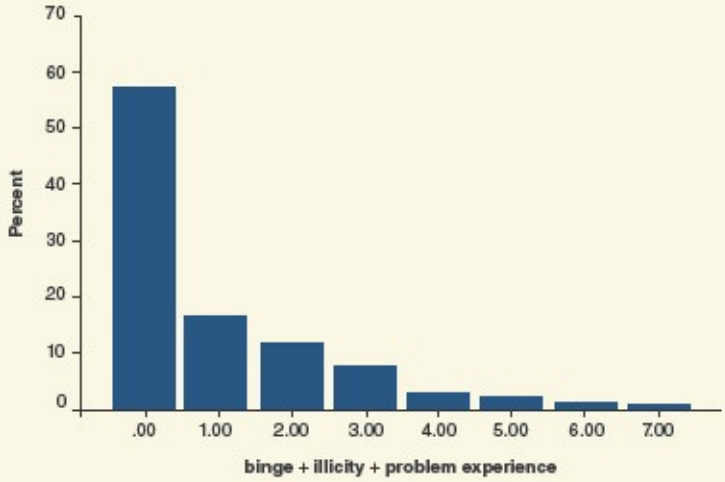
Binge drinking, other forms of alcohol abuse, and illicit drug use create numerous problems on college campuses. Deaths from binge drinking are too common and substance abuse is a factor in as many as two-thirds of on-campus sexual assaults. . . . College presidents now rate alcohol abuse as the number one campus problem . . . many schools have been devising new substance abuse prevention policies and programs. However, in spite of increasing recognition of and knowledge about substance abuse problems at colleges as a whole, little attention has been focused on substance abuse at commuter schools.

FINDINGS

The composite index identifies 27% of respondents as at risk of substance abuse (an index score of 2 or higher). One-quarter reported having smoked or used smokeless tobacco in the past two weeks. 27% of respondents were identified as at risk of substance abuse.

RECOMMENDATIONS

1. Enforce campus rules and regulations about substance use. When possible and where appropriate, communications from campus officials to students should heighten awareness of the UMass–Boston commitment to an alcohol- and drug-free environment.
2. Encourage those students involved in campus alcohol- or drug-related problems or crises to connect with the PRIDE program.
3. Take advantage of widespread student interest in prevention by forming a university-wide council to monitor and stimulate interest in prevention activities.



Source: Schutt, Russell K., Xiaogang Deng, Gerald R. Garrett, Stephanie Hartwell, Sylvia Mignon, Joseph Bebo, Matthew O'Neill, Mary Aruda, Pat Duynstee, Pam DiNapoli, and Helen Reiskin. 1996. Substance use and abuse among UMass Boston students. Unpublished report, Department of Sociology, University of Massachusetts, Boston.

Journal Link

Read an article about race relations that organizes data effectively using charts and tables.

Research in the News

How Much Should Social Scientists Report?

Macro Bertamini and Marcus Munafo discuss a trend emerging in research that puts a premium on shorter scholarly articles. Although “bite size” science enables more work to be published and is supposedly easier to read, Bertamini and Munafo raise a number of concerns. They are also uneasy that articles are only written when researchers are successful. Are scientific failures worth reporting?

1. Is shorter better? What are the advantages and disadvantages of compressing research reports into a few pages or even a paragraph or two? Do social scientists need to balance the type of audience and the complexity of the research in deciding how much to report?
2. “Scientific failures” in this sense mean research projects that result in which no association is found between variables that were hypothesized to be related, or more generally that do not lead to some interesting finding. It is hard to get a paper reporting such null results published because the tendency is to think that the researcher did not “find” anything. Do you see any hazards if social scientists do not publish such “failures”?

News Source: Bertamini, Marco, and Marcus R. Munafo. 2012. The perils of “bite size” science. *New York Times*, January 29: SR 12.

Reverse outlining: Outlining the sections in an already-written draft of a paper or report to improve its organization in the next draft.

Front matter: The section of an applied research report that includes an executive summary, abstract, and table of contents.

Back matter: The section of an applied research report that may include appendixes, tables, and the research instrument(s).

Plagiarism

It may seem depressing to end a book on research methods with a section on plagiarism, but it would be irresponsible to avoid the topic. Of course, you may have a course syllabus detailing instructor or university policies about plagiarism and specifying the penalties for violating that policy, so I'm not simply going to repeat that kind of warning. You probably realize that the practice of selling term papers is revoltingly widespread (a search of "term papers" on Google returned 1,840,000 websites on October 4, 2014); so we're not going to just repeat that academic dishonesty is widespread. Instead, we will use this section to review the concept of plagiarism and to show how that problem connects to the larger issue of the integrity of social research.

You learned in Chapter 3 that maintaining professional integrity—honesty and openness in research procedures and results—is the foundation for ethical research practice. When it comes to research publications and reports, being honest and open means avoiding **plagiarism**—that is, presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment (Hard, Conway, & Moran 2006: 1059).

Now that you are completing this course in research methods, it's time to think about how to do your part to reduce the prevalence of plagiarism. Of course, the first step is to maintain careful procedures for documenting the sources that you rely on for your own research and papers, but you should also think about how best to reduce temptations among others. After all, what people believe about what others do is a strong influence on their own behavior (Hard et al. 2006: 1058).

Reviewing the definition of plagiarism and how your discipline's professional association enforces it is an important first step. This definition and the associated procedures reflect a collective effort to help social scientists maintain standards throughout the discipline (American Sociological Association 1999: 19). The American Sociological Association (ASA)'s (1999) *Code of Ethics* includes an explicit prohibition of plagiarism:

14. Plagiarism

> (a) In publications, presentations, teaching, practice, and service, sociologists explicitly identify, credit, and reference the author when they take data or material verbatim from another person's written work, whether it is published, unpublished, or electronically available.
>
> (b) In their publications, presentations, teaching, practice, and service, sociologists provide acknowledgment of and reference to the use of others' work, even if the work is not quoted verbatim or paraphrased, and they do not present others' work as their own whether it is published, unpublished, or electronically available. (p. 16)

If researchers are motivated by a desire to learn about social relations, to understand how people understand society, and to discover why conflicts arise and how they can be prevented, they will be as concerned with the integrity of their research methods as are those, like yourself, who read and use the results of their research. Throughout *Making Sense of the Social World*, you have been learning how to use research processes and practices that yield valid findings and trustworthy conclusions. Failing to report honestly and openly on the methods used or sources consulted derails progress toward that goal.

Plagiarism: Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.

Conclusion

Good critical skills are essential in evaluating research reports, whether your own or those produced by others. And it is really not just a question of sharpening your knives and going for the jugular. There are *always* weak points in any research, even published research. Being aware of the weaknesses, both in others' studies and in your own, is a major strength in itself. You need to be able to weigh the results of any particular research and to evaluate a study in terms of its contribution to understanding the social world—not in terms of whether it gives a definitive answer for all time, is perfectly controlled, or answers all questions.

Interactive Exercises Link

Reporting Research

This is not to say, however, that "anything goes." Much research lacks one or more of the three legs of validity—measurement validity, causal validity, or generalizability—and contributes more confusion than understanding about the social world. It's true that top scholarly journals maintain very high standards, partly because they have good critics in the review process and distinguished editors who make the final acceptance decisions. But some daily newspapers do a poor job of screening, and research reporting standards in many popular magazines, TV shows, and books are often abysmally poor. Keep your standards high when you read research reports but not so high or so critical that you dismiss studies that make tangible contributions to understanding the social world. And don't be so intimidated by high standards that you shrink from conducting research yourself.

The growth of social science methods from infancy to adolescence, perhaps to young adulthood, ranks as a key intellectual accomplishment of the 20th century. Opinions about the causes and consequences of homelessness no longer need to depend on the scattered impressions of individuals, criminal justice policies can be shaped by systematic evidence of their effectiveness, and changes in the distribution of poverty and wealth in populations can be identified and charted. Employee productivity, neighborhood cohesion, and societal conflict can each be

linked to individual psychological processes and to international economic strains. Systematic researchers looking at truly representative data can make connections and see patterns that no casual observer would ever discern.

Of course, social research methods are only helpful when the researchers are committed and honest. Research methods, like all knowledge, can be used poorly or well, for good purposes or bad, when appropriate or not. A claim that "We're basing this on research!" or "Our statistics prove it!" in itself provides no extra credibility. As you have learned throughout this book, we must first learn which methods were used, how they were applied, and whether final interpretations square with the evidence. But having done all that in good faith, we do emerge from confusion into clarity in our continuing effort to make sense of the social world.

Key Terms

Back matter 331
Front matter 331
Plagiarism 334
Reverse outlining 330

Highlights

- Each research design has strengths and weaknesses. Experimental designs are strong in maximizing causal validity, survey designs maximize generalizability, and qualitative designs maximize authenticity, but tend to be weak in generalizability.
- Research reports should be evaluated systematically, using the review guide in Exhibit 13.2 and considering the interrelations among the design elements.
- Proposal writing should be a time for clarifying the research problem, reviewing the literature, and thinking ahead about the report that will be required. Trade-offs between different design elements should be considered and the potential for mixing methods evaluated.
- Different types of reports typically pose different problems. Authors of student papers must be guided in part by the expectations of their professors. Thesis writers have to meet the requirements of different committee members but can benefit greatly from the areas of expertise represented on a typical thesis committee. Applied researchers are constrained by the expectations of the research sponsor; an advisory committee from the applied setting can help to avoid problems. Journal articles must pass a peer review by other social scientists and often are much improved in the process.
- Research reports should include an introductory statement of the research problem, a literature review, a methodology section, a findings section with pertinent data displays, and a conclusions section that identifies any weaknesses in the research design and points out implications for future research and theorizing. This basic report format should be modified according to the needs of a particular audience.
- All reports should be revised several times and critiqued by others before being presented in final form.
- Plagiarism is too common and should always be rejected.

Student Study Site

Related Titles in Research Methods and Statistics From SAGE

Investigating the Social World: The Process and Practice of Research, Eighth Edition, by Russell K. Schutt

The Practice of Research in Social Work, Third Edition, by Rafael J. Engel and Russell K. Schutt

The Practice of Research in Criminology and Criminal Justice, Fifth Edition, by Ronet K. Bachman and Russell K. Schutt

Fundamentals of Research in Criminology and Criminal Justice, Third Edition, by Ronet K. Bachman and Russell K. Schutt

Designing Surveys: A Guide to Decisions and Procedures, Second Edition, by Ronald F. Czaja and Johnny Blair

A Guide to Qualitative Field Research, Second Edition, by Carol A. Bailey

Adventures in Social Research, Seventh Edition, by Earl Babbie, Fred Halley, William E. Wagner, III, and Jeanne Zaino

Adventures in Criminal Justice Research, Fourth Edition, by Kim Logio, George Dowdall, Earl Babbie, and Fred Halley

Social Statistics for a Diverse Society, Sixth Edition, by Chava Frankfort-Nachmias and Anna Leon-Guerrero

Essentials of Social Statistics for a Diverse Society by Anna Leon-Guerrero and Chava Frankfort-Nachmias

Research Methods in Psychology, Second Edition, by Paul G. Nestor and Russell K. Schutt

Research Methods in Education by Joseph Check and Russell K. Schutt

Fundamentals of Social Work Research, Second Edition, by Rafael J. Engel and Russell K. Schutt

Constructing Social Research by Charles C. Ragin and Lisa M. Amoroso

Using IBM® SPSS® Statistics for Social Statistics and Research Methods, Third Edition, by William E. Wagner, III

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore | Boston

The Student Study Site, available at **edge.sagepub.com/chamblissmssw5e**, includes useful study materials including web exercises with accompanying links, eFlashcards, videos, audio resources, journal articles, and encyclopedia articles, many of which are represented by the media links throughout the text.

Exercises

Discussing Research

1. A good place to start developing your critical skills would be with one of the articles on the study site. Try reading one, and fill in the answers to the article review questions in Exhibit 13.2. Do you agree with our answers to the other questions? Could you add some points to our critique or to the lessons on research design that we drew from these critiques?
2. How firm a foundation do social research methods provide for understanding the social world? Discuss the pro and con arguments, focusing on the variability of social research findings across different social contexts and the difficulty of understanding human subjectivity.

Finding Research

1. Go to the National Science Foundation's Sociology Program website (www.nsf.gov/funding/pgm_summ.jsp?pims_id=5369). What components does the National Science Foundation's Sociology Program look for in a proposed piece of research? Outline a research proposal to study a subject of your choice to be submitted to the National Science Foundation for funding.
2. The National Academy of Sciences wrote a lengthy report on ethics issues in scientific research. Visit the site and read the free executive summary you can obtain (www.nap.edu/catalog.php?record_id=10430). Summarize the information and guidelines in the report.
3. Search a social science journal to find five examples of social science research projects. Briefly describe each. How does each differ in its approach to reporting the research results? To whom do you think the author(s) of each is "reporting" (i.e., who is the audience)? How do you think the predicted audience has helped to shape the author's approach to reporting the results? Be sure to note the source in which you located your five examples.

Critiquing Research

1. A good place to start developing your critical skills would be with Melbin's article that is reviewed in this chapter. Try reading it, and fill in the answers to the article review questions that we did not cover (Exhibit 13.2). Do you agree with our answers to the other questions? Could you add some points to our critique or to the lessons about research designs that we drew from these critiques?
2. Read the journal article "Marital Disruption and Depression in a Community Sample" by Robert Aseltine and Ronald Kessler in the September 1993 issue of *Journal of Health and Social Behavior.* How effective is the article in conveying the design and findings of the research? Could the article's organization be improved at all? Are there bases for disagreement about the interpretation of the findings?
3. Rate four journal articles for overall quality of the research and for effectiveness of the writing and data displays. Discuss how each could have been improved.

Doing Research

1. Call a local social or health service administrator or a criminal justice official, and arrange for an interview. Ask the official about his or her experience with applied research reports and conclusions about the value of social research and the best techniques for reporting to practitioners.
2. Interview a student who has written an independent paper or thesis based on collecting original data. Ask the student to describe her or his experiences while writing the thesis. Review the decisions this student made in designing the research, and ask about the stages of research design, data collection and analysis, and report writing that proved to be difficult.
3. Design a research proposal, following the outline and guidelines presented in this chapter. Focus on a research question that you could study on campus or in your local community.

Ethics Questions

1. Plagiarism is no joke. What are the regulations on plagiarism in class papers at your school? What do you think the ideal policy would be? Should this policy account for cultural differences in teaching practices and learning styles? Do you think this ideal policy is likely to be implemented? Why or why not? Based on your experiences, do you believe that most student plagiarism is the result of misunderstanding about proper citation practices, or is it the result of dishonesty? Do you think that students who plagiarize while in school are less likely to be honest as social researchers?
2. Most journals now require full disclosure of funding sources, as well as paid consulting and other business relationships. Should researchers publishing in social science journals also be required to fully disclose all sources of funding, including receipt of payment for research done as a consultant? Should full disclosure of all previous funding sources be required in each published article? Write a short justification of the regulations you propose.

Video Interview Questions

1. Listen to Schutt's interview for Chapter 13 at edge.sagepub.com/chamblissmssw5e.
2. What were the primary research findings?
3. What changes did the Women's Health Network implement in light of research findings?

Glossary

Alternate-forms reliability:
A procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

Anomalous:
Unexpected patterns in data that do not seem to fit the theory being proposed.

Anonymity:
Provided by research in which no identifying information is recorded that could be used to link respondents to their responses.

Archival data:
Written or visual records, not produced by the researcher.

Association:
A criterion for establishing a causal relationship between two variables: Variation in one variable is empirically related to variation in another variable.

Availability sampling:
Sampling in which elements are selected on the basis of convenience.

Back matter:
The section of an applied research report that may include appendixes, tables, and the research instrument(s).

Bar chart:
A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces.

Base number (*N*):

The total number of cases in a distribution.

Before-and-after design:
A quasi-experimental design consisting of several before-after comparisons involving the same variables but no comparison group.

***Belmont Report*:**
Report in 1979 of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research stipulating three basic ethical principles for the protection of human subjects: respect for persons, beneficence, and justice.

Beneficence:
Minimizing possible harms and maximizing benefits.

Bias:
Sampling bias occurs when some population characteristics are over- or underrepresented in the sample because of particular features of the method of selecting the sample.

Big Data:
Data produced or accessible in computer-readable form that is produced by people, available to social scientists, and manageable with today's computers.

Bimodal:
A distribution in which two nonadjacent categories have about the same number of cases and these categories have more cases than any others.

Case-oriented research:
Research that focuses attention on the nation or other unit as a whole.

Causal effect:
The finding that change in one variable leads to change in another variable, *ceteris paribus* (other things being equal). *Example:* Individuals arrested for domestic assault tend to commit fewer subsequent assaults than similar individuals who are accused in the same circumstances but are not arrested.

Causal (internal) validity:
Exists when a conclusion that A leads to, or results in, B is correct.

Census:

Research in which information is obtained through responses from or information about all available members of an entire population.

Central tendency:

The most common value (for variables measured at the nominal level) or the value around which cases tend to center (for a qualitative variable).

Certificate of Confidentiality:

Document issued by the National Institutes of Health to protect researchers from being legally required to disclose confidential information.

Ceteris paribus:

Latin phrase meaning “other things being equal.”

Chi-square:

An inferential statistic used to test hypotheses about relationships between two or more variables in a cross-tabulation.

Closed-ended (fixed-choice) question:

A survey question that provides preformatted response choices for the respondent to circle or check.

Cluster:

A naturally occurring, mixed aggregate of elements of the population.

Cluster sampling:

Sampling in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Cognitive interview:

A technique for evaluating questions in which researchers ask people test questions, and then probe with follow-up questions to learn how they understood the question and what their answers mean.

Cohort:

Individuals or groups with a common starting point.

Cohort design:

A longitudinal study in which data are collected at two or more points in time from individuals in a cohort.

Comparison groups:

In an experiment, groups that have been exposed to different treatments or values of the independent variable (e.g., a control group and an experimental group).

Compensatory rivalry (John Henry effect):

A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied the treatment and modify their efforts by way of compensation.

Complete observation:

A role in participant observation in which the researcher does not participate in group activities and is publicly defined as a researcher.

Complete (covert) participation:

A role in field research in which the researcher does not reveal his or her identity as a researcher to those who are observed.

Computer-assisted personal interview (CAPI):

A personal interview in which the laptop computer is used to display interview questions and to process responses that the interviewer types in, as well as to check that these responses fall within allowed ranges.

Computer-assisted qualitative data analysis:

Analysis of textual, aural, or pictorial data using a special computer program that facilitates searching and coding text.

Concept:

A mental image that summarizes a set of similar observations, feelings, or ideas.

Conceptualization:

The process of specifying what we mean by a term. In deductive research, conceptualization helps translate portions of an abstract theory into testable hypotheses involving specific variables. In inductive research, conceptualization is an important part of the process used to make sense of related observations.

Confidentiality:
Provided by research in which identifying information that could be used to link respondents to their responses is available only to designated research personnel for specific research needs.

Constant:
A number that has a fixed value in a given situation; a characteristic or value that does not change.

Construct validity:
The type of validity that is established by showing that a measure is related to other measures as specified in a theory.

Contamination:
A source of causal invalidity that occurs when either the experimental or the comparison group is aware of the other group and is influenced in the posttest as a result.

Content analysis:
A research method for systematically analyzing and making inferences from text.

Context:
The larger set of interrelated circumstances in which a particular outcome should be understood.

Context effects:
In survey research, refers to the influence that earlier questions may have on how subsequent questions are answered.

Contingent question:
A question that is asked of only a subset of survey respondents.

Contrived observation:
Observations of situations in which the researcher has deliberately intervened.

Control group:
A comparison group that receives no treatment.

Cost-benefit analysis:
A type of evaluation research that compares program costs with the economic value of program benefits.

Cost-effectiveness analysis:
A type of evaluation research that compares program costs with actual program outcomes.

Cover letter:
The letter sent with a mailed questionnaire that explains the survey's purpose and auspices and encourages the respondent to participate.

Criterion validity:
The type of validity that is established by comparing the scores obtained on the measure being validated to those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

Cross-population generalizability (external validity):
Exists when findings about one group, population, or setting hold true for other groups, populations, or settings.

Cross-sectional research design:
A study in which data are collected at only one point in time.

Cross-tabulation (crosstab):
In the simplest case, a bivariate (two-variable) distribution showing the distribution of one variable for each category of another variable; can also be elaborated using three or more variables.

Data cleaning:
The process of checking data for errors after the data have been entered in a computer file.

Debriefing:
A researcher's informing subjects after an experiment about the experiment's purposes and methods and evaluating subjects' personal reactions to the experiment.

Deductive research:
The type of research in which a specific expectation is deduced from a

general premise and is then tested.

Demoralization:
A type of contamination in experimental and quasi-experimental designs that occurs when control group members feel that they have been left out of some valuable treatment, performing worse than expected as a result.

Dependent variable:
A variable that is hypothesized to vary depending on or under the influence of another variable.

Descriptive research:
Research in which social phenomena are defined and described.

Descriptive statistics:
Statistics used to describe the distribution of and relationship among variables.

Differential attrition (mortality):
A problem that occurs in experiments when comparison groups become different because subjects in one group are more likely to drop out for various reasons compared with subjects in the other group(s).

Direction of association:
A pattern in a relationship between two variables—that is, the value of a variable tends to change consistently in relation to change in the other variable. The direction of association can be either positive or negative.

Disproportionate stratified sampling:
Sampling in which elements are selected from strata in proportions different from those that appear in the population.

Distribution of benefits:
An ethical issue about how much researchers can influence the benefits subjects receive as part of the treatment being studied in a field experiment.

Double-barreled question:
A single survey question that actually asks two questions but allows only one answer.

Double-blind procedure:

An experimental method in which neither the subjects nor the staff delivering experimental treatments know which subjects are getting the treatment.

Double negative:

A question or statement that contains two negatives, which can muddy the meaning of the question.

Ecological fallacy:

An error in reasoning in which conclusions about individual-level processes are drawn from group-level data.

Effect of external events:

*See***History effect**.

Efficiency analysis:

A type of evaluation research that compares program costs with program effects. It can be either a cost-benefit analysis or a cost-effectiveness analysis.

Elaboration analysis:

The process of introducing a third variable into an analysis to better understand —to elaborate—the bivariate (two-variable) relationship under consideration; additional control variables also can be introduced.

Electronic survey:

A survey that is sent and answered by computer, either through e-mail or on the web.

Elements:

The individual members of the population whose characteristics are to be measured.

E-mail survey:

A survey that is sent and answered through e-mail.

Emic focus:

Representing a setting with the participants' terms.

Endogenous change:

A source of causal invalidity that occurs when natural developments or changes in the subjects (independent of the experimental treatment itself) account for some or all of the observed change from the pretest to the posttest.

Ethnography:

The study and systematic recording of human cultures.

Ethnomethodology:

A qualitative research method focused on the way that participants in a social setting create and sustain a sense of reality.

Etic focus:

Representing a setting with the researcher's terms.

Evaluability assessment:

A type of evaluation research conducted to determine whether it is feasible to evaluate a program's effects within the available time and resources.

Evaluation research:

Research that describes or identifies the impact of social policies and programs.

Event-structure analysis:

A systematic method of developing a causal diagram showing the structure of action underlying some chronology of events; the result is an idiographic causal explanation.

Exhaustive:

Every case can be classified as having at least one attribute (or value) for the variable.

Expectancies of experiment staff (self-fulfilling prophecy):

A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects results from the positive expectancies of the staff who are delivering the treatment, rather than to the treatment itself.

Experimental group:

In an experiment, the group of subjects that receives the treatment or experimental manipulation.

Explanatory research:
Seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in another phenomenon.

Exploratory research:
Seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them.

Ex post facto control group design:
A nonexperimental design in which comparison groups are selected after the treatment, program, or other variation in the independent variable has occurred.

Extraneous variable:
A variable that influences both the independent and dependent variables to create a spurious association between them that disappears when the extraneous variable is controlled.

Face validity:
The type of validity that exists when an inspection of items used to measure a concept suggests that they are appropriate "on their face."

Federal Policy for the Protection of Human Subjects:
Federal regulations codifying basic principles for conducting research on human subjects; used as the basis for professional organizations' guidelines.

Feedback:
Information about service delivery system outputs, outcomes, or operations that is available to any program inputs.

Fence-sitters:
Survey respondents who see themselves as being neutral on an issue and choose a middle (neutral) response that is offered.

Field experiment:
An experimental study conducted in a real-world setting.

Field notes:

Notes that describe what has been observed, heard, or otherwise experienced in a participant observation study. These notes usually are written after the observational session.

Field research:

Research in which natural social processes are studied as they happen and left relatively undisturbed.

Filter question:

A survey question used to identify a subset of respondents who then are asked other questions.

Fixed-choice question:

*See***Closed-ended question.**

Floaters:

Survey respondents who provide an opinion on a topic in response to a closed-ended question that does not include a "Don't know" option but who will choose "Don't know" if it is available.

Focus groups:

A qualitative method that involves unstructured group interviews in which the focus group leader actively encourages discussion among participants on the topics of interest.

Formative evaluation:

Process evaluation that is used to shape and refine program operations.

Frequency distribution:

Numerical display showing the number of cases, and usually the percentage of cases (the relative frequencies), corresponding to each value or group of values of a variable.

Frequency polygon:

A graphic for quantitative variables in which a continuous line connects data points representing the variable's distribution.

Front matter:

The section of an applied research report that includes an executive

summary, abstract, and table of contents.

Gamma:
A measure of association that is sometimes used in cross-tabular analysis.

Gatekeeper:
A person in a field setting who can grant researchers access to the setting.

Generalizability:
Exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify; it is the extent to which a study can inform us about persons, places, or events that were not directly studied.

Grounded theory:
Systematic theory developed inductively, based on observations that are summarized into conceptual categories, reevaluated in the research setting, and gradually refined and linked to other conceptual categories.

Group-administered survey:
A survey that is completed by individual respondents who are assembled in a group.

Group unit of analysis:
A unit of analysis in which groups are the source of data and the focus of conclusions.

Hawthorne effect:
A type of contamination in experimental and quasi-experimental designs that occurs when members of the treatment group change relative to the dependent variable because their participation in the study makes them feel special.

Health Insurance Portability and Accountability Act (HIPAA):
A U.S. federal law passed in 1996 that guarantees, among other things, specified privacy rights for medical patients, in particular those in research settings.

Histogram:
A graphic for quantitative variables in which the variable's distribution is

displayed with adjacent bars.

History effect (effect of external events):
Events external to the study that influence posttest scores, resulting in causal invalidity.

Holistic research:
Research concerned with the context in which events occurred and the interrelations between different events and processes.

Hypothesis:
A tentative statement about empirical reality involving a relationship between two or more variables.

Illogical reasoning:
The premature jumping to conclusions or arguing on the basis of invalid assumptions.

Impact analysis (impact evaluation or summative evaluation):
Evaluation research that answers these questions: Did the program work? Did it have the intended result?

Independent variables:
A variable that is hypothesized to cause, or lead to, variation in another variable.

Index:
A composite measure based on summing, averaging, or otherwise combining the responses to multiple questions that are intended to measure the same concept.

Individual unit of analysis:
A unit of analysis in which individuals are the source of data and the focus of conclusions.

Inductive reasoning:
The type of reasoning that moves from the specific to the general.

Inductive research:
The type of research in which general conclusions are drawn from specific

data.

Inferential statistics:
Statistics used to estimate how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

In-person interview:
A survey in which an interviewer questions respondents face-to-face and record their answers.

Inputs:
Resources, raw materials, clients, and staff that go into a program.

Institutional review board (IRB):
A group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to subjects.

Instrument decay:
The deterioration over time of a measurement instrument, resulting in increasingly inaccurate results.

Integrative approaches (to evaluation):
An orientation to evaluation research that expects researchers to respond to the concerns of people involved with the program stakeholders, as well as to the standards and goals of the social scientific community.

Intensive (depth) interviewing:
A qualitative method that involves open-ended, relatively unstructured questioning in which the interviewer seeks in-depth information on the interviewee's feelings, experiences, and perceptions.

Interactive voice response (IVR):
A survey in which respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software.

Interitem reliability (internal consistency):

An approach that calculates reliability based on the correlation between multiple items used to measure a single concept.

Interobserver reliability:
When similar measurements are obtained by different observers rating the same persons, events, or places.

Interpretive questions:
Questions included in a questionnaire or interview schedule to help explain answers to other important questions.

Interquartile range:
The range in a distribution between the end of the lst quartile and the beginning of the 3rd quartile.

Interval level of measurement:
A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.

Interview schedule:
A survey instrument containing the questions asked by the interviewer in an in-person or phone survey.

John Henry effect:
*See***Compensatory rivalry**.

Jottings:
Brief notes written in the field about highlights of an observation period.

Justice:
As used in human research ethics discussions, distributing benefits and risks of research fairly.

Key informant:
An insider who is willing and able to provide a field researcher with superior access and information, including answers to questions that arise during the research.

Level of measurement:

The mathematical precision with which the values of a variable can be expressed. The nominal level of measurement, which is qualitative, has no mathematical interpretation; the quantitative levels of measurement—ordinal, interval, and ratio—are progressively more precise mathematically.

Longitudinal research design:
A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.

Mailed (self-administered) survey:
A survey involving a mailed questionnaire to be completed by the respondent.

Matching:
A procedure for equating the characteristics of individuals in different comparison groups in an experiment. Matching can be done on either an individual or an aggregate basis. For individual matching, individuals who are similar in key characteristics are paired before assignment, and then the two members of each pair are assigned to the two groups. For aggregate matching, groups chosen for comparison are similar in the distribution of key characteristics.

Matrix:
A chart used to condense qualitative data into simple categories and provide a multidimensional summary that will facilitate subsequent, more intensive analysis.

Mean:
The arithmetic, or weighted, average computed by adding up the value of all the cases and dividing by the total number of cases.

Measurement validity:
Exists when an indicator measures what we think it measures.

Measure of association:
A type of descriptive statistic that summarizes the strength of an association.

Mechanism:
A discernible process that creates a causal connection between two

variables.

Median:
The position average, or the point, that divides a distribution in half (the 50th percentile).

Mode (probability average):
The most frequent value in a distribution.

Mortality:
*See***Differential attrition**.

Multiple group before-and-after design:
A type of quasi-experimental design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.

Mutually exclusive:
A variable's attributes (or values) are mutually exclusive when every case can be classified as having only one attribute (or value).

Narrative analysis:
A form of qualitative analysis in which the analyst focuses on how respondents impose order on the flow of experience in their lives and so make sense of events and actions in which they have participated.

Narrative explanation:
An explanation that involves developing a narrative of events and processes that indicate a chain of causes and effects.

Needs assessment:
A type of evaluation research that attempts to determine the needs of some population that might be met with a social program.

Netnography (cyberethnography or virtual ethnography):
The use of ethnographic methods to study online communities.

Ngrams:
Frequency graphs produced by Google's database of all words printed in more than one third of the world's books over time (with coverage still

expanding).

Nominal level of measurement:
Variables whose values have no mathematical interpretation; they vary in kind or quality but not amount.

Nonequivalent control group design:
A quasi-experimental design in which there are experimental and comparison groups that are designated before the treatment occurs but are not created by random assignment.

Nonprobability sampling methods:
Sampling methods in which the probability of selection of population elements is unknown.

Nonspuriousness:
A criterion for establishing a causal relation between two variables; when a relationship between two variables is not caused by variation in a third variable.

Normal distribution:
A symmetric distribution shaped like a bell and centered around the population mean, with the number of cases tapering off in a predictable pattern on both sides of the mean.

Nuremberg war crime trials:
Trials held in Nuremberg, Germany, in the years following World War II, in which the former leaders of Nazi Germany were charged with war crimes and crimes against humanity; frequently considered the first trials for people accused of genocide.

Obedience experiments (Milgram's):
A series of famous experiments conducted during the 1960s by Stanley Milgram, a psychologist from Yale University, testing subjects' willingness to cause pain to another person if instructed to do so.

Office for Protection From Research Risks, National Institutes of Health:
Federal agency that monitors institutional review boards (IRBs).

Omnibus survey:

A survey that covers a range of topics of interest to different social scientists.

Open-ended question:
A survey question to which the respondents reply in their own words, either by writing or by talking.

Operation:
A procedure for identifying or indicating the value of cases on a variable.

Operationalization:
The process of specifying the operations that will indicate the value of cases on a variable.

Oral history:
Data collected through intensive interviews with participants in past events.

Ordinal level of measurement:
A measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.

Outcomes:
The impact of the program process on the cases processed.

Outlier:
An exceptionally high or low value in a distribution.

Outputs:
The services delivered or new products produced by the program process.

Overgeneralization:
Occurs when we unjustifiably conclude that what is true for some cases is true for all cases.

Panel design:
A longitudinal study in which data are collected from the same individuals —the panel—at two or more points in time.

Participant observation:

A qualitative method for gathering data that involves developing a sustained relationship with people while they go about their normal activities.

Percentage:
The relative frequency, computed by dividing the frequency of cases in a particular category by the total number of cases and multiplying by 100.

Periodicity:
A sequence of elements (in a list to be sampled) that varies in some regular, periodic pattern.

Phone survey:
A survey in which interviewers question respondents over the phone and record their answers.

Physical traces:
Either the erosion or the accumulation of physical substances that can be used as evidence of activity. For instance, footprints in snow indicate that someone has walked there

Placebo effect:
A source of treatment misidentification that can occur when subjects receive a treatment that they consider likely to be beneficial and improve as a result of the expectation rather than of the treatment itself.

Plagiarism:
Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.

Population:
The entire set of individuals or other entities to which study findings are to be generalized.

Posttest:
In experimental research, the measurement of an outcome (dependent) variable after an experimental intervention or after a presumed independent variable has changed for some other reason. The posttest is exactly the same "test" as the pretest, but it is administered at a different time.

Pretest:

In experimental research, the measurement of an outcome (dependent) variable before an experimental intervention or change in a presumed independent variable for some other reason. The pretest is exactly the same "test" as the posttest, but it is administered at a different time.

Prison simulation study (Zimbardo's):

Famous study from the early 1970s, organized by Stanford psychologist Philip Zimbardo, demonstrating the willingness of average college students quickly to become harsh disciplinarians when put in the role of (simulated) prison guards over other students; usually interpreted as demonstrating an easy human readiness to become cruel.

Probability average:

*See***Mode**.

Probability of selection:

The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all the elements of a population, the probability that any particular element will be selected is 1.0. If half the elements in the population are sampled on the basis of chance (say, by tossing a coin), the probability of selection for each element is one half, or 0.5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

Probability sampling method:

A sampling method that relies on a random, or chance, selection method so that the probability of selection of population elements is known.

Process analysis:

A research design in which periodic measures are taken to determine whether a treatment is being delivered as planned, usually in a field experiment.

Process evaluation:

Evaluation research that investigates the process of service delivery.

Program process:

The complete treatment or service delivered by the program.

Program theory:

A descriptive or prescriptive model of how a program operates and produces effects.

Progressive focusing:
The process by which a qualitative analyst interacts with the data and gradually refines his or her focus.

Proportionate stratified sampling:
Sampling method in which elements are selected from strata in exact proportion to their representation in the population.

Purposive sampling:
A nonprobability sampling method in which elements are selected for a purpose, usually because of their unique position.

Qualitative data analysis:
Techniques used to search and code textual, aural, and pictorial data and to explore relationships among the resulting categories.

Qualitative methods:
Methods, such as participant observation, intensive interviewing, and focus groups, that are designed to capture social life as participants experience it rather than in categories the researcher predetermines. These methods typically involve exploratory research questions, inductive reasoning, an orientation to social context, and a focus on human subjectivity and the meanings participants attach to events and to their lives.

Quantitative data analysis:
Statistical techniques used to describe and analyze variation in quantitative measures.

Quartiles:
The points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases.

Quasi-experimental design:
A research design in which there is a comparison group that is comparable to the experimental group in critical ways but subjects are not randomly assigned to the comparison and experimental groups.

Questionnaire:
A survey instrument containing the questions in a self-administered survey.

Quota sampling:
A nonprobability sampling method in which elements are selected to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Random assignment (randomization):
A procedure by which each experimental subject is placed in a group randomly.

Random digit dialing (RDD):
The random dialing, by a machine, of numbers within designated phone prefixes, which creates a random sample for phone surveys.

Random number table:
A table containing lists of numbers that are ordered solely on the basis of chance; it is used for drawing a random sample.

Random sampling:
A method of sampling that relies on a random, or chance, selection method so that every element of the sampling frame has a known probability of being selected.

Range:
The true upper limit in a distribution minus the true lower limit (or the highest rounded value minus the lowest rounded value, plus 1).

Ratio level of measurement:
A measurement of a variable in which the numbers indicating the variable's values represent fixed measuring units *and* an absolute zero point.

Reactive effects:
The changes in an individual or group behavior that result from being observed or otherwise studied.

Reactive methods:
When the people being studied know they are being studied, and so may modify their answers or even the behavior being studied itself.

Reductionist fallacy (reductionism):
An error in reasoning that occurs when incorrect conclusions about group-level processes are based on individual-level data.

Regression effect:
A source of causal validity that occurs when subjects chosen because of their extreme scores on a dependent variable become less extreme on a posttest as a result of mathematical necessity, rather than the treatment.

Reliability:
A measurement procedure yields consistent scores when the phenomenon being measured is not changing.

Repeated cross-sectional studies:
*See***Trend designs**.

Repeated measures panel design:
A quasi-experimental design consisting of several pretest and posttest observations of the same group.

Representative sample:
A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Research circle:
A diagram of the elements of the research process, including theories, hypotheses, data collection, and data analysis.

Resistance to change:
The reluctance to change our ideas in light of new information.

Respect for persons:
In human subjects ethics discussions, treating persons as autonomous agents and protecting those with diminished autonomy.

Reverse outlining:

Outlining the sections in an already-written draft of a paper or report to improve its organization in the next draft.

Sample:
A subset of a population that is used to study the population as a whole.

Sample generalizability:
Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

Sampling frame:
A list of all elements or other units containing the elements in a population.

Sampling interval:
The number of cases between one sampled case and another in a systematic random sample.

Sampling units:
Units listed at each stage of a multistage sampling design.

Saturation point:
The point at which subject selection is ended in intensive interviewing because new interviews seem to yield little additional information.

Scale:
A composite measure based on combining the responses to multiple questions pertaining to a common concept after these questions are differentially weighted, such that questions judged on some basis to be more important for the underlying concept contribute more to the composite score.

Science:
A set of logical, systematic, documented methods for investigating nature and natural processes; the knowledge produced by these investigations.

Secondary data:
Previously collected data that are used in a new analysis.

Secondary data analysis:
The method of using preexisting data in a different way or to answer a

different research question than intended by those who collected the data.

Selection bias:
A source of internal (causal) invalidity that occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.

Selective (inaccurate) observation:
Choosing to look only at things that are in line with our preferences or beliefs.

Self-fulfilling prophecy:
See Expectancies of experiment staff.

Serendipitous:
Unexpected patterns in data, which stimulate new ideas or theoretical approaches.

Simple random sampling:
A method of sampling in which every sample element is selected purely on the basis of chance through a random process.

Skewness:
The extent to which cases are clustered more at one or the other end of the distribution of a quantitative variable rather than in a symmetric pattern around its center. Skew can be positive (a right skew), with the number of cases tapering off in the positive direction, or negative (a left skew), with the number of cases tapering off in the negative direction.

Skip pattern:
The unique combination of questions created in a survey by filter questions and contingent questions.

Snowball sampling:
A method of sampling in which sample elements are selected as successive informants or interviewees identify them.

Social research question:
A question about the social world that is answered through the collection and analysis of firsthand, verifiable, empirical data.

Social science:
The use of scientific methods to investigate individuals, societies, and social processes; the knowledge produced by these investigations.

Social science approaches (to evaluation):
An orientation to evaluation research that expects researchers to emphasize the importance of researcher expertise and maintenance of autonomy from program stakeholders.

Split-halves reliability:
Reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.

Spurious:
Nature of a presumed relationship between two variables that actually results from variation in a third variable.

Stakeholder approaches (to evaluation):
An orientation to evaluation research that expects researchers to be responsive primarily to the people involved with the program.

Stakeholders:
Individuals and groups who have some basis of concern with the program.

Standard deviation:
The square root of the average squared deviation of each case from the mean.

Statistic:
A numerical description of some feature of a variable or variables in a sample from a larger population.

Statistical significance:
The mathematical likelihood that an association is not the result of chance, judged by a criterion the analyst sets.

Stratified random sampling:
A method of sampling in which sample elements are selected separately from population strata that the researcher identifies in advance.

Summative evaluation:
*See***Impact analysis**.

Survey research:
Research in which information is collected from a sample of individuals through their responses to a set of standardized questions.

Systematic random sampling:
A method of sampling in which sample elements are selected from a list or from sequential files, with every *n*th element being selected after the first element is selected randomly.

Tacit knowledge:
In field research, a credible sense of understanding of social processes that reflects the researcher's awareness of participants' actions, as well as their words, and of what they fail to state, feel deeply, and take for granted.

Target population:
A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.

***Tearoom Trade*:**
Book by Laud Humphreys investigating the social background of men who engage in homosexual behavior in public facilities; controversially, he did not obtain informed consent from his subjects.

Test–retest reliability:
A measurement showing that measures of a phenomenon at two points in time are highly correlated, if the phenomenon has not changed or has changed only as much as the phenomenon itself.

Theoretical sampling:
A sampling method recommended for field researchers by Glaser and Strauss (1967). A theoretical sample is drawn in a sequential fashion, with settings or individuals selected for study as earlier observations or interviews indicate that these settings or individuals are influential.

Theory:
A logically interrelated set of propositions about empirical reality.

Theory-driven evaluation:
A program evaluation guided by a theory that specifies the process by which the program has an effect.

Time order:
A criterion for establishing a causal relationship between two variables: The variation in the presumed cause (the independent variable) must occur before the variation in the presumed effect (the dependent variable).

Time series design:
A quasi-experimental design consisting of many pretest and posttest observations of the same group.

Treatment misidentification:
A problem that occurs in an experiment when not the treatment itself, but rather some unknown or unidentified intervening process is causing the outcome.

Trend (repeated cross-sectional) design:
A longitudinal study in which data are collected at two or more points in time from different samples of the same population.

Triangulation:
The use of multiple methods to study one research question.

True experiment:
Experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation. Outcomes are measured in a posttest.

Tuskegee syphilis study:
Research study conducted by a branch of the U.S. government, lasting for roughly 50 years (ending in the 1970s), in which a sample of African American men diagnosed with syphilis were deliberately left untreated, without their knowledge, to learn about the lifetime course of the disease.

Unimodal:
A distribution of a variable in which only one value is the most frequent.

Units of analysis:
The entities being studied, whose behavior is to be understood.

Unobtrusive measure:
A measurement based on physical traces or other data that are collected without the knowledge or participation of the individuals or groups that generated the data.

Validity:
The state that exists when statements or conclusions about empirical reality are correct.

Variability:
The extent to which cases are spread out through the distribution or clustered around just one value.

Variable:
A characteristic or property that can vary (take on different values or attributes).

Variable-oriented research:
Research that focuses attention on variables representing particular aspects of the cases studied and then examines the relations between these variables across sets of cases.

Variance:
A statistic that measures the variability of a distribution as the average squared deviation of each case from the mean.

Web survey:
A survey that is accessed and responded to on the World Wide Web.

Appendix A Finding Information

Elizabeth Schneider, MLS

Russell K. Schutt, PhD

All research is conducted to “find information” in some sense, but the focus of this section is more specifically about finding information to inform a central research project. This has often been termed *searching the literature*, but the popularity of the World Wide Web for finding information requires that we broaden our focus beyond the traditional search of the published literature. It may sound trite, but we do indeed live in an “information age,” with an unprecedented amount of information of many types available to us with relatively little effort. Learning how to locate and use that information efficiently has become a prerequisite for social science.

Searching the Literature

It is most important to search the literature before we begin a research study. A good literature review may reveal that the research problem already has been adequately investigated, it may highlight particular aspects of the research problem most in need of further investigation, or it may suggest that the planned research design is not appropriate for the problem chosen. A good literature review can highlight the strong and weak points of related theories. When we review previous research about our research question, we may learn about weaknesses in our measures, complexities in our research problem, and possible difficulties in data collection. The more of these problems that can be considered before, rather than after, data are collected, the better the final research product will be. Even when the rush to “find out” what people think or are doing creates pressure to just go out and ask or observe, it is important to take the time to search the literature and try to reap the benefit of prior investigations.

But the social science literature is not just a source for guidance at the start of an investigation. During a study, questions will arise that can be answered by careful reading of earlier research. After data collection has ceased, reviewing the literature can help you develop new insights into patterns in the data. Research articles published since a project began may suggest new hypotheses or questions to explore.

The best way of searching the literature will be determined partly by what library and bibliographic resources are available to you, but a brief review of some basic procedures and alternative strategies will help you get started on a productive search.

Preparing the Search

You should formulate a research question before you begin the search, although the question may change after you begin. Identify the question's parts and subparts and any related issues that you think might play an important role in the research. List the authors of relevant studies you are aware of, possible keywords that might specify the subject for your search, and perhaps the most important journals that you are concerned with checking. For example, if your research question is "What is the effect of informal social control on crime?" you might consider searching the literature electronically for studies that mentioned "informal social control" and "crime" or "crime rate" or "violence" and "arrest." If you are concerned with more specific aspects of this question, you should also include the relevant words in your list, such as *family* or *community policing* or even *Northeast*.

Conducting the Search

Now you are ready to begin searching the literature. You should check for relevant books in your library and perhaps in the other college libraries in your area. This usually means conducting a search of an online catalog using a list of subject terms. But most scientific research is published in journal articles so that research results can quickly be disseminated to other scientists. The primary focus of your search must therefore be the journal literature. Fortunately, much of the journal literature can be identified online, without leaving your personal computer, and an increasing number of published journal articles can be downloaded directly to your own computer (depending on your particular access privileges). But just because there's a lot available online doesn't mean that you need to find it all. Keep in mind that your goal is to find reports of prior research investigations; this means that you should focus on scholarly journals that choose articles for publication after they have been reviewed by other social scientists—"refereed journals." Newspaper and magazine articles just won't do, although you may find some that raise important issues or even that summarize social science research investigations.

The social science literature should be consulted at both the beginning and the end of an investigation. Even while an investigation is in progress, consultations with the literature may help resolve methodological problems or facilitate supplementary explorations. As with any part of the research process, the method you use will affect the quality of your results. You should try to ensure that your search method includes each of the following steps:

Specify your research question.

Your research question should not be so broad that hundreds of articles are judged relevant, or so narrow that you miss important literature. "Is informal social control effective?" is probably too broad. "Does informal social control reduce rates of burglary in large cities?" is probably too narrow. "Is informal social control more effective in reducing crime rates than policing?" provides about the right level of specificity.

Identify appropriate bibliographic databases to search.

Your school library may subscribe to Sociological Abstracts or SocINDEX and either of these similar databases of the sociological literature may meet your needs. You can limit your searches in these databases to articles written in English, articles that have been peer reviewed and so are likely to be of higher quality, and to articles in journals that your library owns. However, if you are studying a question about social factors in illness you should also search in MEDLINE or the slightly more comprehensive PubMed, the databases for searching the medical literature maintained by the National Library of Medicine. If your focus is on mental health, you'll also want to include a search in the psychological abstracts, with PsycARTICLES (or PsycINFO, if that is what your library offers). Searching in a database such as Academic OneFile or Google Scholar will retrieve article abstracts across disciplines, but it will be important to review your results very carefully to ensure that the articles you focus on are appropriate for a sociological research paper. To find articles across the social sciences that have referred to a previous publication, such as Lawrence Sherman and Richard Berk's study of the police response to domestic violence, the Social Science Citation Index (SSCI) will be helpful. SSCI has a unique "citation searching" feature that allows you to look up articles or books and see who else has cited them in their work. This is an excellent and efficient way to assemble a number of references that are highly relevant to your research and to find out which articles and books have had the biggest impact in a field. Unfortunately, some college libraries do not subscribe to SSCI, but if you have access to it, you should consider using it to make sure that you develop the strongest possible literature review for your topic.

Choose a search technology.

For most purposes, an online bibliographic database that references the published journal literature will be all you need to find the relevant social science research literature. However, searches for more obscure topics or very recent literature may require that you also search websites or bibliographies of relevant books. You will also need to search websites when you need to learn about current debate about particular social issues or you are investigating current social programs.

Create a tentative list of search terms.

List the parts and subparts of your research question and any related issues that you think are important: "informal social control," "policing," "influences on

crime rates," and perhaps "community cohesion and crime." List the authors of relevant studies. Specify the most important journals that deal with your topic.

Narrow your search.

The sheer number of references you find can be a problem. For example, searching for peer reviewed journal articles on "social capital" in October 2014 resulted in 4,108 citations in Sociological Abstracts and 5,386 in SocINDEX to peer reviewed articles written in English scholarly journals. Depending on the database you are working with and the purposes of your search, you may want to limit your search to English language publications, to journal articles rather than conference papers or dissertations (both of which are more difficult to acquire), and to materials published in recent years. You should give most attention to articles published in the leading journals in the field. Your professor can help you identify them.

Refine your search.

Learn as you go. If your search yields too many citations, try specifying the search terms more precisely. If you have not found much literature, try using more general terms. Whatever terms you search on first, don't consider your search complete until you have tried several different approaches and have seen how many articles you find. A search for "domestic violence" in SocINDEX on October 10, 2014, yielded 3,880 abstracts for peer-reviewed journal articles in English; by adding "effects" OR "influences" as required search terms the number of hits dropped to 405. A good rule is to cast a net with your search terms that is wide enough to catch most of the relevant articles but not so wide that it identifies many useless citations. In any case, if you are searching a popular topic, you will need to spend a fair amount of time whittling down the list of citations.

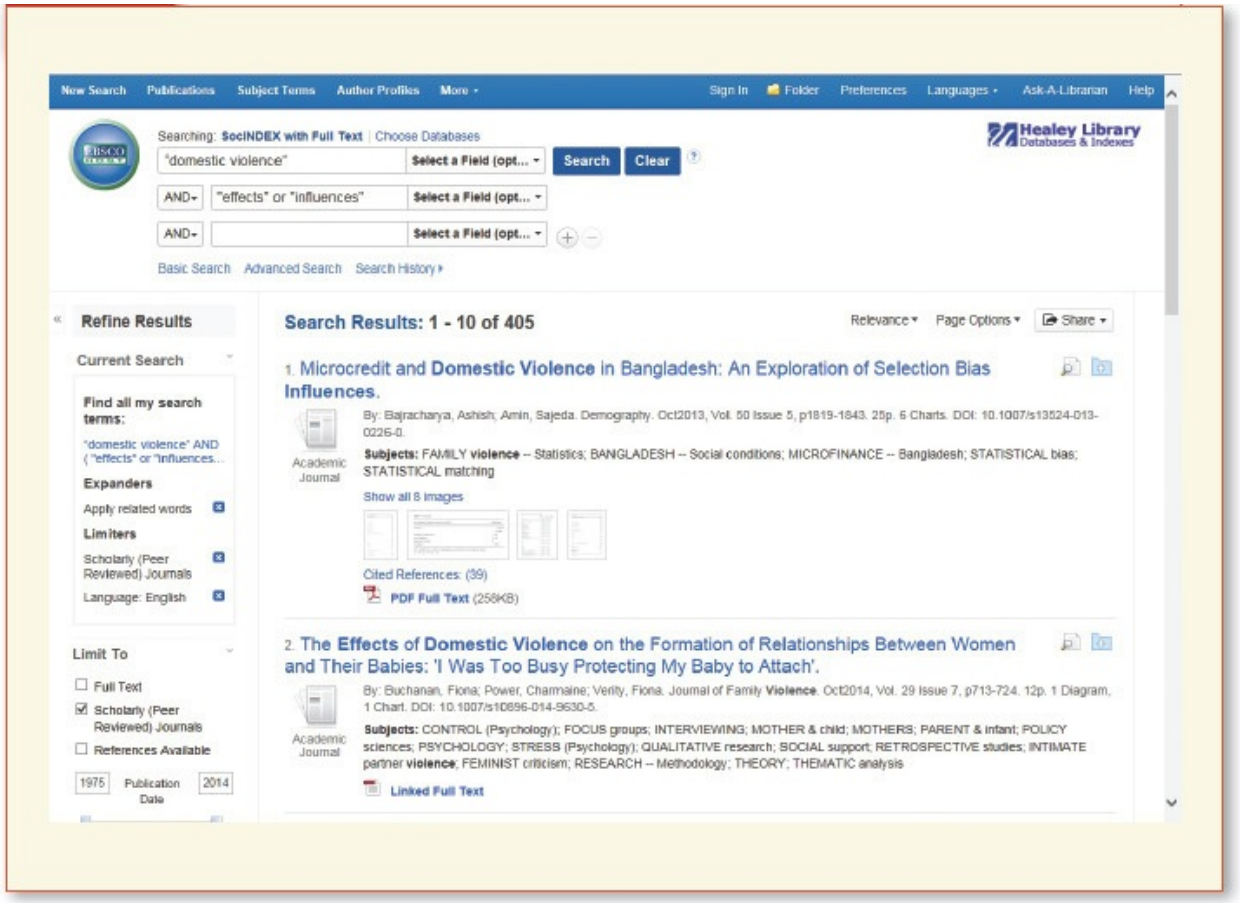
Use Boolean search logic.

It's often a good idea to narrow your search by requiring that abstracts contain combinations of words or phrases that include more of the specifics of your research question. Using the Boolean connector AND allows you to do this, whereas using the connector OR allows you to find abstracts containing different words that mean the same thing. Exhibit A.1 provides an example.

Use appropriate subject descriptors.

Once you have found an article that you consider appropriate, look at the "Subject Terms" field in the citation (see Exhibit A.2). You can then redo your search after requiring that the articles be classified with some or all of these descriptor terms.

Exhibit A.1 Use of Boolean Connectors in a Literature Search


New Search
Publications
Subject Terms
Author Profiles
More
Sign In
Folder
Preferences
Languages
Ask A Librarian
Help
Searching: SocINDEX with Full Text | Choose Databases
Healey Library
Databases & Indexes
"domestic violence"
Select a Field (opt...
Search
Clear
AND
"effects" or "influences"
Basic Search
Advanced Search
Search History
Refine Results
Search Results: 1 - 10 of 405
Relevance
Page Options
Share
Current Search
Find all my search terms:
"domestic violence" AND ("effects" or "influences...
Expanders
Apply related words
Limiters
Scholarly (Peer Reviewed) Journals
Language: English
Limit To
Full Text
Scholarly (Peer Reviewed) Journals
References Available
1975
Publication Date
2014
1. Microcredit and Domestic Violence in Bangladesh: An Exploration of Selection Bias Influences.
Academic Journal
By: Bajracharya, Ashish; Amin, Sajeda. Demography. Oct2013, Vol. 50 Issue 5, p1819-1843. 25p. 6 Charts. DOI: 10.1007/s13524-013-0226-0.
Subjects: FAMILY violence -- Statistics; BANGLADESH -- Social conditions; MICROFINANCE -- Bangladesh; STATISTICAL bias; STATISTICAL matching
Show all 8 images
Cited References: (39)
PDF Full Text (258KB)
2. The Effects of Domestic Violence on the Formation of Relationships Between Women and Their Babies: 'I Was Too Busy Protecting My Baby to Attach'.
Academic Journal
By: Buchanan, Fiona; Power, Charmaine; Verity, Fiona. Journal of Family Violence. Oct2014, Vol. 29 Issue 7, p713-724. 12p. 1 Diagram, 1 Chart. DOI: 10.1007/s10896-014-9630-5.
Subjects: CONTROL (Psychology); FOCUS groups; INTERVIEWING; MOTHER & child; MOTHERS; PARENT & infant; POLICY sciences; PSYCHOLOGY; STRESS (Psychology); QUALITATIVE research; SOCIAL support; RETROSPECTIVE studies; INTIMATE partner violence; FEMINIST criticism; RESEARCH -- Methodology; THEORY; THEMATIC analysis
Linked Full Text

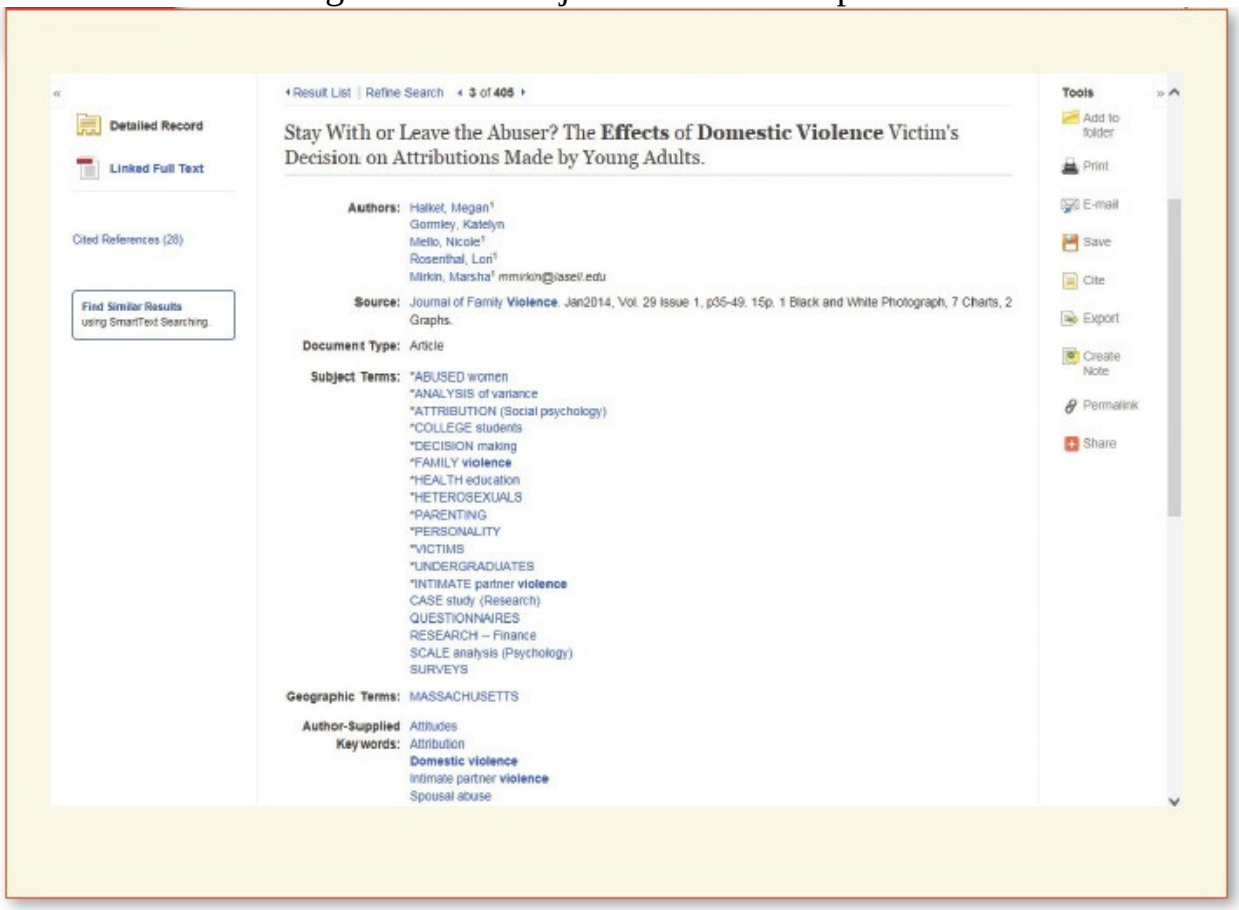
Check the results.

Read the titles and abstracts you have found and identify the articles that appear to be most relevant. If possible, click on these article titles and generate a list of their references. See if you find more articles that are relevant to your research question but that you have missed so far. You will be surprised (we always are) at how many important articles your initial online search missed.

Read the articles.

Now it is time to find the full text of the articles of interest. If you're lucky, many of the articles will be available to patrons of your library in online versions. If so, you'll be able to link to the full text just by clicking on a "full text" link. But many journals or specific issues of some journals will only be available in print, so you'll have to find them in your library (or order a copy through interlibrary loan). You may be tempted to write a "review" of the literature based on reading the abstracts or using only those articles available online, but you will be selling yourself short. Many crucial details about methods, findings, and theoretical implications will be found only in the body of the article and some important articles will not be available online. To understand, critique, and really learn from previous research studies, you must read the important articles, no matter how you have to retrieve them. But if you can't obtain the full text of an article, you'll just have to leave it out of your literature review and bibliography—reading the abstract just isn't enough.

Exhibit A.2 Checking Standard Subject Matter Descriptors

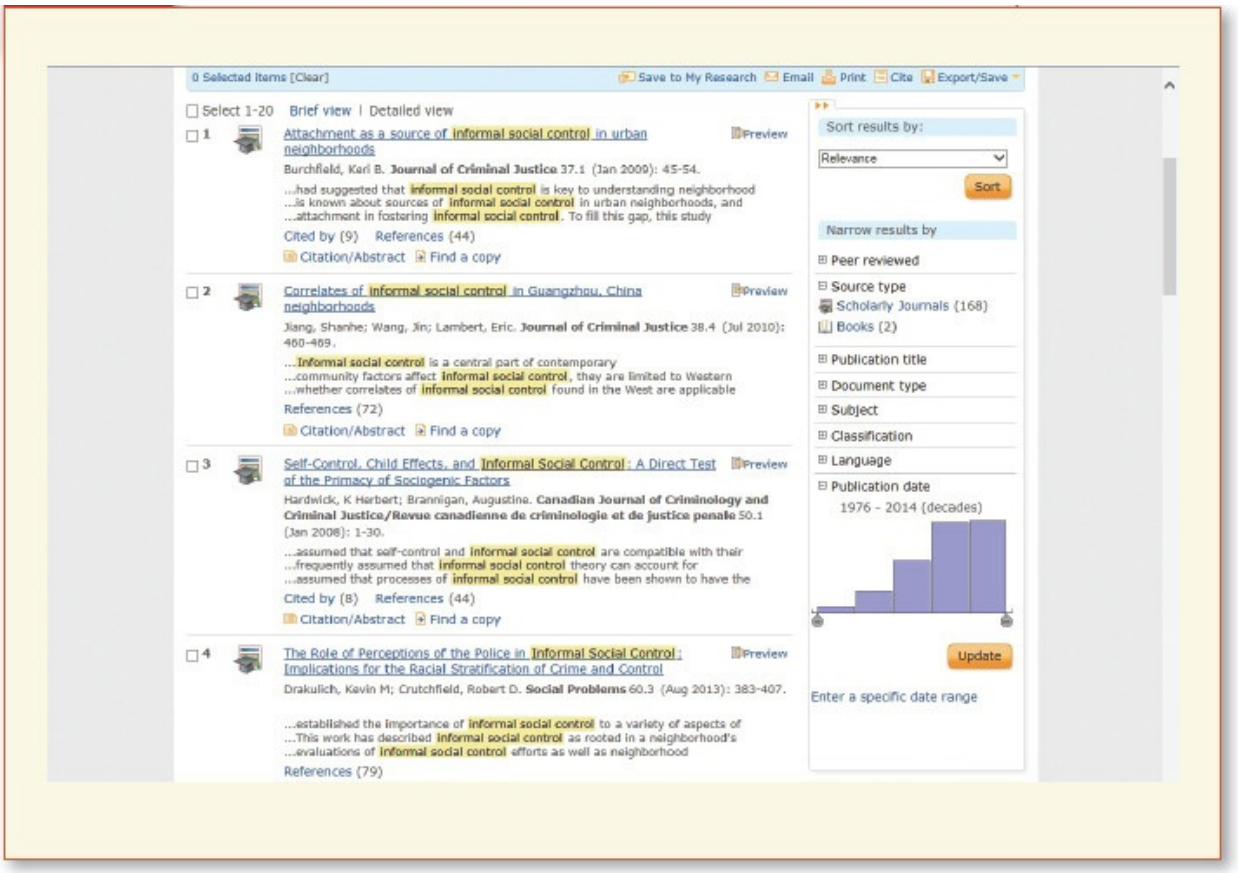


Write the review.

If you have done your job well, you will now have more than enough literature as background for your own research unless it is on a very obscure topic (see Exhibit A.3). (Of course, ultimately your search will be limited by the library holdings you have access to and by the time you have to order or find copies of journal articles, conference papers, and perhaps dissertations that you can't obtain online.) At this point, your main concern is to construct a coherent framework in which to develop your research question, drawing as many lessons as you can from previous research. You can use the literature to identify a useful theory and hypotheses to be reexamined, to find inadequately studied specific research questions, to explicate the disputes about your research question, to summarize the major findings of prior research, and to suggest appropriate methods of investigation.

Be sure to take notes on each article you read, organizing your notes into standard sections: theory, methods, findings, conclusions. In any case, write the literature review so that it contributes to your study in some concrete way; don't feel compelled to discuss an article just because you have read it. Be judicious. You are conducting only one study of one issue; it will only obscure the value of your study if you try to relate it to every tangential point in related research.

Exhibit A.3 A Search in Sociological Abstracts on "Informal Social Control"


0 Selected Items [Clear]
Save to My Research
Email
Print
Cite
Export/Save
Select 1-20
Brief view | Detailed view
1
Attachment as a source of informal social control in urban neighborhoods
Preview
Burchfield, Keri B. Journal of Criminal Justice 37.1 (Jan 2009): 45-54.
...had suggested that informal social control is key to understanding neighborhood
...is known about sources of informal social control in urban neighborhoods, and
...attachment in fostering informal social control. To fill this gap, this study
Cited by (9) References (44)
Citation/Abstract Find a copy
2
Correlates of informal social control in Guangzhou, China neighborhoods
Preview
Jiang, Shanhe; Wang, Jin; Lambert, Eric. Journal of Criminal Justice 38.4 (Jul 2010): 460-469.
...Informal social control is a central part of contemporary
...community factors affect informal social control, they are limited to Western
...whether correlates of informal social control found in the West are applicable
References (72)
Citation/Abstract Find a copy
3
Self-Control, Child Effects, and Informal Social Control : A Direct Test of the Primacy of Sociogenic Factors
Preview
Hardwick, K Herbert; Brannigan, Augustine. Canadian Journal of Criminology and Criminal Justice/Revue canadienne de criminologie et de justice penale 50.1 (Jan 2008): 1-30.
...assumed that self-control and informal social control are compatible with their
...frequently assumed that informal social control theory can account for
...assumed that processes of informal social control have been shown to have the
Cited by (8) References (44)
Citation/Abstract Find a copy
4
The Role of Perceptions of the Police in Informal Social Control : Implications for the Racial Stratification of Crime and Control
Preview
Drakulich, Kevin M; Crutchfield, Robert D. Social Problems 60.3 (Aug 2013): 383-407.
...established the importance of informal social control to a variety of aspects of
...This work has described informal social control as rooted in a neighborhood's
...evaluations of informal social control efforts as well as neighborhood
References (79)
Sort results by:
Relevance
Sort
Narrow results by
Peer reviewed
Source type
Scholarly Journals (168)
Books (2)
Publication title
Document type
Subject
Classification
Language
Publication date
1976 - 2014 (decades)
Update
Enter a specific date range

Continue to search.

Don’t think of searching the literature as a one-time-only venture—something that you leave behind as you move on to your *real* research. You may encounter new questions or unanticipated problems as you conduct your research or as you burrow deeper into the literature. Searching the literature again to determine what others have found in response to these questions or what steps they have taken to resolve these problems can yield substantial improvements in your own research. There is so much literature on so many topics that it often is not possible to figure out in advance every subject you should search the literature for or what type of search will be most beneficial.

Another reason to make searching the literature an ongoing project is that the literature is always growing. During the course of one research study, whether it takes only one semester or several years, new findings will be published and relevant questions will be debated. Staying attuned to the literature and checking it at least when you are writing up your findings may save your study from being outdated. Of course, this does not make life any easier for researchers. For

example, one of your authors was registered for a time with a service that every week sent citations of new journal articles on homelessness to his electronic mailbox. Most were not very important, and even looking over the abstracts for between 5 and 15 new articles each week is quite a chore—that's part of the price we pay for living in the information age!

Refer to a good book for even more specific guidance about literature searching. Arlene Fink's (2005) *Conducting Research Literature Reviews: From the Internet to Paper* is an excellent guide.

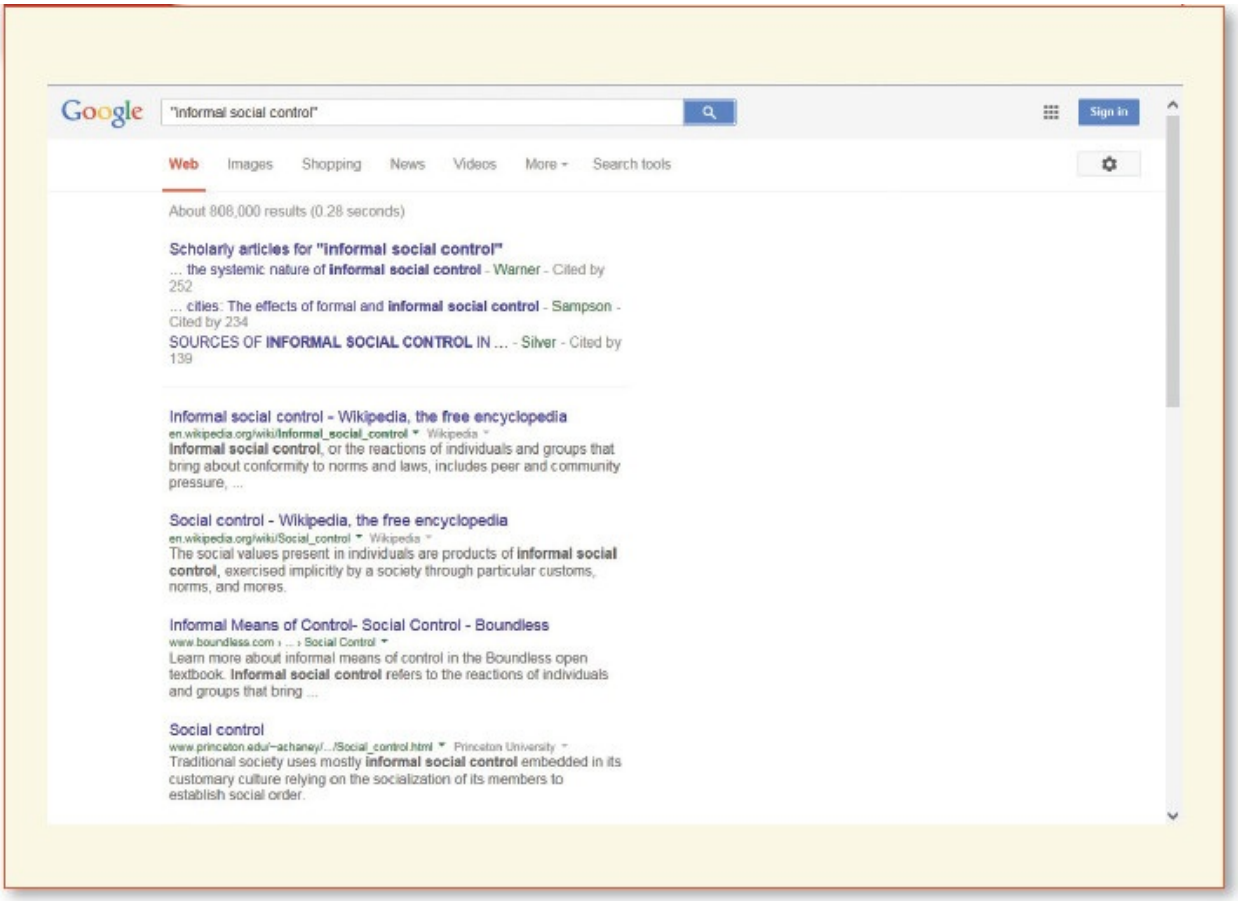
Searching the Web

The World Wide Web provides access to vast amounts of information of many different sorts (O'Dochartaigh 2002). You can search the holdings of other libraries and download the complete text of government reports, some conference papers, many books, and newspaper articles. You can find policies of local governments, descriptions of individual social scientists and particular research projects, and postings of advocacy groups. It's also hard to avoid finding a lot of information in which you have no interest, such as commercial advertisements, third-grade homework assignments, or college course syllabi. In October 2014, there were 4.31 billion pages on the web (http://www.worldwidewebsize.com/).

After you are connected to the web with a browser such as Microsoft Internet Explorer or Google Chrome or Mozilla Firefox, you can use three basic strategies for finding information: direct addressing—typing in the address, or URL, of a specific site; browsing—reviewing online lists of websites; and searching—the most common approach. Google is currently the most popular search engine for searching the web. For some purposes, you will need to use only one strategy; for other purposes, you will want to use all three. End-of-chapter web exercises and the SAGE Edge Study Site for this text both provide many URLs relevant to social science research.

Exhibit A.4 illustrates the first problem that you may encounter when searching the web: the sheer quantity of resources that are available. It is a much bigger problem than when searching bibliographic databases. On the web, less is usually more. Limit your inspection of websites to the first few pages that turn up in your list (they're ranked by relevance). See what those first pages contain, and then try to narrow your search by including some additional terms. Putting quotation marks around a phrase that you want to search will also help limit your search—searching for "informal social control" on Google (on October 10, 2014) produced 808,000 sites, compared with the 2,380,000 sites retrieved when we omitted the quotes, so Google searched "informal" *and* "social" *and* "control."

Exhibit A.4 Google Search Results for "Informal Social Control"


Google
"informal social control"
Sign in
Web
Images
Shopping
News
Videos
More
Search tools
About 808,000 results (0.28 seconds)
Scholarly articles for "informal social control"
... the systemic nature of informal social control - Warner - Cited by 252
... cities: The effects of formal and informal social control - Sampson - Cited by 234
SOURCES OF INFORMAL SOCIAL CONTROL IN ... - Silver - Cited by 139
Informal social control - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Informal_social_control
Informal social control, or the reactions of individuals and groups that bring about conformity to norms and laws, includes peer and community pressure, ...
Social control - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Social_control
The social values present in individuals are products of informal social control, exercised implicitly by a society through particular customs, norms, and mores.
Informal Means of Control- Social Control - Boundless
www.boundless.com › ... › Social Control
Learn more about informal means of control in the Boundless open textbook. Informal social control refers to the reactions of individuals and groups that bring ...
Social control
www.princeton.edu/~achaney/.../Social_control.html
Traditional society uses mostly informal social control embedded in its customary culture relying on the socialization of its members to establish social order.

Remember the following warnings when you conduct searches on the web:

- *Clarify your goals.* Before you begin the search, jot down the terms that you think you need to search for as well as a statement of what you want to accomplish with your search. This will help to ensure that you have a sense of what to look for and what to ignore.
- *Quality is not guaranteed.* Anyone can post almost anything, so the accuracy and adequacy of the information you find are always suspect. There's no journal editor or librarian to evaluate quality and relevance. You need to anticipate the different sources of information available on the web and to decide whether it is appropriate to use each of them for specific purposes. The sources you will find include:
    - *Books*—Google is scanning the text of books that are out-of-print or no longer protected by copyright. In 2013, the total number of books scanned was over 30 million (out of more than 130 million books in the world) (http://en.wikipedia.org/wiki/Google_Books). When you search in Google Books, you will retrieve the pages in books that use the cited terms.

    - *Newspaper articles*—These can range from local newspapers such as the *Chicago Tribune* to national newspapers such as the *New York Times*. Access to articles in these newspapers may be limited to subscribers.
    - *Government policies*—You can find government policies and publications ranging from those done at the city or town level to those written by foreign governments.
    - *Presented papers*—You may find the complete text of a formal presentation that was given at a meeting or conference.
    - *Classroom lecture notes and outlines; listings from college catalogs*—These are pretty straightforward.
    - *Commercial advertisements*—Advertising abounds on the web, and it is especially prolific on search engine pages. Your search engine will even retrieve ads from the web and list them as results of your search! The boundaries between academic, nonprofit, and commercial information have become very porous, so you can't let your guard down.
- *Anticipate change.* Websites that are not maintained by stable organizations can come and go very quickly. Any search will result in attempts to link to some URLs that no longer exist.
- *One size does not fit all.* Different search engines use different procedures for indexing websites. Some attempt to be all-inclusive, whereas others aim to be selective. As a result, you can get different results from different search engines (such as Google or Bing) even though you are searching for exactly the same terms.
- *Be concerned about generalizability.* You might be tempted to characterize police department policies by summarizing the documents you find at police department websites. But how many police departments are there? How many have posted their policies on the web? Are these policies representative of all police departments? To answer all these questions, you would have to conduct a research project just on the websites themselves.
- *Evaluate the sites.* There's a lot of stuff out there; so how do you know what's good? Some websites contain excellent advice and pointers on how to differentiate the good from the bad. You can find one example at http://olinuris.library.cornell.edu/ref/research/webeval.html.
- *Avoid web addiction.* Another danger of the extraordinary quantity of information available on the web is that one search will lead to another and to another and. . . . There are always more possibilities to explore and one more interesting source to check. Establish boundaries of time and effort to

avoid the risk of losing all sense of proportion.

- *Cite your sources.* Using text or images from web sources without attribution is plagiarism. It is the same as copying someone else's work from a book or article and pretending that it is your own. Record the web address (URL), the name of the information provider, and the date on which you obtain material from the site. Include this information in a footnote to the material that you use in a paper.

References

Abbott, Andrew. 1994. History and sociology: The lost synthesis. Pp. 77–112 in Engaging the past: The uses of history across the social sciences, ed. Eric H. Monkkonen. Durham, NC: Duke University Press.

Adler, Patricia A. 1993. Wheeling and dealing: An ethnography of an upper-level drug dealing and smuggling community, 2nd ed. New York: Columbia University Press.

Adler, Patricia A., and Peter Adler. 2000. Intense loyalty in organizations: A case study of college athletics. Pp. 31–50 in Qualitative studies of organizations, ed. John Van Maanen. Thousand Oaks, CA: Sage.

Adorno, Theodor W., Nevitt Sanford, Else Frenkel-Brunswik, and Daniel Levinson. 1950. The authoritarian personality. New York: Harper.

Aiden, Erez, and Jean-Baptiste Michel. 2013. Uncharted: Big data as a lens on human culture. New York: Riverhead Books.

Alfred, Randall. 1976. The church of Satan. Pp. 180–202 in The new religious consciousness, eds. Charles Glock and Robert Bellah. Berkeley: University of California Press.

Altheide, David L., and John M. Johnson. 1994. Criteria for assessing interpretive validity in qualitative research. Pp. 485–499 in Handbook of qualitative research, eds. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Altman, Lawrence K. 1987. U.S. and France end rift on AIDS. New York Times, April 1. From http://www.nytimes.com (accessed May 28, 2007).

Ambwani, Suman, and Jaine Strauss. 2007. Love thyself before loving others? A qualitative and quantitative analysis of gender differences in body image and romantic love. Sex Roles 56: 13–21.

American Sociological Association (ASA). 1997. Code of ethics. Washington, DC: American Sociological Association.

Anderson, Elijah. 2000. Code of the street: Decency, violence, and the moral life of the inner city, reprint ed. New York: W. W. Norton.

Anderson, Elijah. 2003. Jelly's place: An ethnographic memoir (distinguished lecture). Symbolic Interaction 26(2): 217–237.

Anspach, Renee R. 1991. Everyday methods for assessing organizational effectiveness. Social Problems 38(February): 1–19.

Aronson, Elliot, and Judson Mills. 1959. The effect of severity of initiation on liking for a group. Journal of Abnormal and Social Psychology 59(September): 177–181.

Arwood, Tracy, and Sangeeta Panicker. 2007. Assessing risk in social and behavioral sciences. From Collaborative Institutional Training Initiative website: https://www.citiprogram.org/members/learners/ (accessed June 5, 2008).

Aseltine Jr., Robert H., and Ronald C. Kessler. 1993. Marital disruption and depression in a community sample. Journal of Health and Social Behavior 34 (September): 237–251.

Associated Press. 2000. Researchers fear privacy breaches with online research. From http://www.pi.edu/News/TechNews/000919/onlineresearch.html (accessed September 28, 2008).

Ayhan, H. Öztaş. 2001. Gender bias in agricultural surveys statistics by gender: Measures to reduce gender bias in agricultural surveys. International Statistical Review 69: 447–460.

Ayres, Ian, and Peter Siegelman. 1995. Race and gender discrimination in bargaining for a new car. American Economic Review 85(3): 304–321.

Ayres, Ian, Frederick E. Vars, and Nasser Zakariya. 2005. To insure prejudice: Racial disparities in taxicab tipping. Yale Law Journal 114: 1613.

Bachman, Ronet, and Russell K. Schutt. 2007. The practice of research in criminology and criminal justice, 3rd ed. Thousand Oaks, CA: Sage.

Bail, Christopher A. 2008. The configuration of symbolic boundaries against immigrants in Europe. American Sociological Review 73: 37–59.

Bakalar, Nicholas. 2010. Child abuse investigations didn't reduce risk, a study finds. New York Times, October 12: D3.

Banks, J. A. 1972. Historical sociology and the study of population. Pp. 55–70 in Population and Social Change, eds. D. V. Glass and Roger Revelle. London: Edward Arnold.

Baumrind, Diana. 1964. Some thoughts on ethics of research: After reading Milgram's "Behavioral study of obedience." American Psychologist 19: 421–423.

Baumrind, Diana. 1985. Research using intentional deception: Ethical issues revisited. American Psychologist 40: 165–174.

Becker, Howard S. 1958. Problems of inference and proof in participant observation. American Sociological Review 23: 652–660.

Becker, Howard S. 1963. Outsiders: Studies in the sociology of deviance. New York: Free Press.

*** Becker, Howard S. 1986. Writing for social scientists. Chicago: University of Chicago Press. [This can be ordered directly from the American Sociological Association, 1722 N St. NW, Washington, DC 20036, (202) 833–3410.]

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. Habits of the heart: Individualism and commitment in American life. New York: Harper & Row.

Bendix, Reinhard. 1956. Work and authority in industry: Ideologies of management in the course of industrialization. Berkeley: University of California Press.

Bendix, Reinhard. 1962. Max Weber: An intellectual portrait. Garden City, NY: Doubleday/Anchor.

Benzecry, Claudio. 2011. The opera fanatic: Ethnography of an obsession. Chicago: University of Chicago Press.

Bertamini, Marco, and Marcus R. Munafo. 2012. The perils of “bite size” science. New York Times, January 29: SR 12.

Bierman, Alex. 2012. Functional limitations and psychological distress: Marital status as moderator. Society and Mental Health 2(1): 35–52.

Binder, Arnold, and James W. Meeker. 1993. Implications of the failure to replicate the Minneapolis experimental findings. American Sociological Review 58: 886–888.

Birkeland, Sarah, Erin Murphy-Graham, and Carol Weiss. 2005. Good reasons for ignoring good evaluation: The case of the Drug Abuse Resistance Education (D.A.R.E.) program. Evaluation and Program Planning 28: 247–256.

Boase, Jeffrey, John B. Horrigan, Barry Wellman, and Lee Rainie. 2006. The strength of Internet ties: The Internet and email aid users in maintaining their social networks and provide pathways to help when people face big decisions. Washington, DC: Pew Internet & American Life Project.

Bogdewic, Stephan P. 1999. Participant observation. Pp. 47–70 in Doing qualitative research, 2nd ed., eds. Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Bohn, Angela, Christian Buchta, Kurt Hornik, and Patrick Mair. 2014. Making friends and communicating on Facebook: Implications for the access to social capital. Social Networks 37: 29–41.

Bollen, Kenneth A., Barbara Entwisle, and Arthur S. Alderson. 1993. Macrocomparative Research Methods. Annual Review of Sociology 19: 321–351.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams. 1995. The craft of research. Chicago: University of Chicago Press.

Boruch, Robert F. 1997. Randomized experiments for planning and evaluation: A practical guide. Thousand Oaks, CA: Sage.

Bramel, Dana, and Ronal Friend. 1981. Hawthorne, the myth of the docile worker, and class bias in psychology. American Psychologist 38 (September): 867–878.

Brewer, John, and Albert Hunter. 1989. Multimethod research: A synthesis of

styles. Newbury Park, CA: Sage.

Brooks, Clem, and Jeff Manza. 2006. Social policy responsiveness in developed democracies. American Sociological Review 71: 474–494.

Brown v. Board of Education (Brown I), 347 U.S. 483 (1954).

Butler, Declan. 2013. When Google got flu wrong. Nature 494 (February 14): 155–156.

Butterfield, Fox. 1996a. After 10 years, juvenile crime begins to drop. New York Times, August 9: A1, A25.

Butterfield, Fox. 1996b. Gun violence may be subsiding, studies find. New York Times, October 14: A10.

Butterfield, Fox. 2000. As murder rates edge up, concern, but few answers. New York Times, June 18: A12.

Buzawa, Eve S., and Carl G. Buzawa, eds. 1996. Do arrests and restraining orders work? Thousand Oaks, CA: Sage.

Campbell, Donald T., and Julian C. Stanley. 1966. Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campbell, Donald T., and M. Jean Russo. 1999. Social experimentation. Thousand Oaks, CA: Sage.

Campbell, Richard T. 1992. Longitudinal research. Pp. 1146–1158 in Encyclopedia of sociology, eds. Edgar F. Borgatta and Marie L. Borgatta. New York: Macmillan.

Campbell, Wilson. 2002. A statement from the Governmental Accounting Standards Board and Performance Measurement Staff. American Society for Public Administration. From http://64.91.242.87/cap/forum_statement.html (accessed July 20, 2002).

Cava, Anita, Reid Cushman, and Kenneth Goodman. 2007. HIPAA and human subjects research. From the University of Miami website: http://researchedu.med.miami.edu/x36.xml (accessed September 28, 2008).

Cave, Emma, and Soren Holm. 2003. Milgram and Tuskegee: Paradigm research projects in bioethics. Health Care Analysis 11: 27–40.

Center for Survey Research, University of Massachusetts at Boston. 1987. Methodology: Designing good survey questions. Newsletter, April 3.

Chambliss, Daniel F. 1988. Champions: The making of Olympic swimmers. New York: Morrow.

Chambliss, Daniel F. 1989. The mundanity of excellence: An ethnographic report on stratification and Olympic swimmers. Sociological Theory 7(1): 70–86.

Chambliss, Daniel F. 1996. Beyond caring: Hospitals, nurses, and the social organization of ethics. Chicago: University of Chicago Press.

Chambliss, Daniel F., and Christopher G. Takacs. 2014. How college works. Cambridge, MA: Harvard University Press.

Chen, Huey-Tsyh. 1990. Theory-driven evaluations. Newbury Park, CA: Sage.

Chen, Huey-Tsyh, and Peter H. Rossi. 1987. The theory-driven approach to validity. Evaluation and Program Planning 10: 95–103.

Coffey, Amanda, and Paul Atkinson. 1996. Making sense of qualitative data: Complementary research strategies. Thousand Oaks, CA: Sage.

Cohen, Alison K., and Benjamin W. Chaffee. 2012. The relationship between adolescents' civic knowledge, civic attitude, and civic behavior and their self-reported future likelihood of voting. Education, Citizenship and Social Justice 8(1): 43–57.

Cohen, Gary E., and Barbara A. Kerr. 1998. Computer-mediated counseling: An empirical study of a new mental health treatment. Computers in Human Services 15: 13–26.

Cohen, Susan G., and Gerald E. Ledford Jr. 1994. The effectiveness of self-managing teams: A quasi-experiment. Human Relations 47: 13–43.

Coleman, James S., and Thomas Hoffer. 1987. Public and private high schools: The impact of communities. New York: Basic Books.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. High school achievement: Public, Catholic, and private schools compared. New York: Basic Books.

Collins, Randall. 1975. Conflict sociology: Toward an explanatory science. New York: Academic Press.

Collins, Randall. 2009. Violence: A micro-sociological theory. Princeton, NJ: Princeton University Press.

Connor, Jennie, Andrew Gray, and Kypros Kypri. 2010. Drinking history, current drinking and problematic sexual experiences among university students. Australian and New Zealand Journal of Public Health 34: 487–494.

Converse, Jean M. 1984. Attitude measurement in psychology and sociology: The early years. Pp. 3–40 in Surveying subjective phenomena, vol. 2, eds. Charles F. Turner and Elizabeth Martin. New York: Russell Sage Foundation.

Cook, Gareth. 2011. TV's sleeper effect. Boston Globe, October 30.

Cook, Thomas D., and Donald T. Campbell. 1979. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.

Coontz, Stephanie. 1997. The way we really are: Coming to terms with America's changing families. New York: Basic Books.

Cooper, Kathleen B., and Michael D. Gallagher. 2004. A nation online: Entering the broadband age. Washington, DC: Economics and Statistics Administration and National Telecommunications and Information Administration, U.S. Department of Commerce.

Core Institute. 1994. Core alcohol and drug survey: Long form. Carbondale, IL: Fund for the Improvement of Postsecondary Education (FIPSE) Core Analysis Grantee Group, Core Institute, Student Health Programs, Southern Illinois University.

Costner, Herbert L. 1989. The validity of conclusions in evaluation research: A further development of Chen and Rossi's theory-driven approach. Evaluation and Program Planning 12: 345–353.

Couper, Mick P. 2000. Web surveys: A review of issues and approaches. Public Opinion Quarterly 64: 464–494.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. Web survey design and administration. Public Opinion Quarterly 65: 230–253.

Couper, Mick P., and Peter V. Miller. 2008. Web survey methods: Introduction. Public Opinion Quarterly 72: 831–835.

Couper, Mick P., Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls II, and James M. O'Reilly (Eds.). 1998. Computer-assisted survey information collection. New York: Wiley.

Cowan, Gloria. 2002. Content analysis of visual materials. Pp. 345–368 in Handbook for conducting research on human sexuality, eds. Michael W. Wiederman and Bernard E. Whitley Jr. London: Lawrence Erlbaum.

Cross, Harry, Genevieve Kenney, Jane Mell, and Wendy Zimmerman. 1990. Employer hiring practices: Differential treatment of Hispanic and Anglo job seekers. Washington, DC: Urban Institute Press.

Dannefer, W. Dale, and Russell K. Schutt. 1982. Race and juvenile justice processing in court and police agencies. American Journal of Sociology 87 (March): 1113–1132.

D.A.R.E. 2008. The "New" D.A.R.E. Program. From www.dare.com/newdare.asp (accessed May 31, 2008).

Davidoff, Steven M. 2012. Seeking critical mass of gender equality in the boardroom. New York Times, September 12: B5.

Davies, Philip, Anthony Petrosino, and Iain Chalmers. 1999. Report and papers from the exploratory meeting for the Campbell Collaboration. London: School of Public Policy, University College.

Davis, James A. 1985. The logic of causal order. Sage University Paper Series on Quantitative Applications in the Social Sciences (series no. 07- 055). Beverly Hills, CA: Sage.

Davis, James A., and Tom W. Smith. 1992. The NORC General Social Survey: A user's guide. Newbury Park, CA: Sage.

Dawes, Robyn. 1995. How do you formulate a testable exciting hypothesis? Pp. 93–96 in How to write a successful research grant application: A guide for social and behavioral scientists, eds. Willo Pequegnat and Ellen Stover. New York: Plenum Press.

Decker, Scott H., and Barrik Van Winkle. 1996. Life in the gang: Family, friends, and violence. New York: Cambridge University Press.

de Leeuw, Edith. 2008. Self-administered questionnaires and standardized interviews. Pp. 311–327 in The SAGE handbook of social research methods, eds. Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Demos, John. 1998. History beyond data bits. New York Times, December 30: A23.

Dentler, Robert A. 2002. Practicing sociology: Selected fields. Westport, CT: Praeger.

Denzin, Norman K. 2002. The interpretive process. Pp. 349–368 in The qualitative researcher's companion, eds. A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage.

Denzin, Norman K., and Yvonna S. Lincoln. 1994. Introduction: Entering the field of qualitative research. Pp. 1–17 in Handbook of qualitative research, eds. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Denzin, Norman K., and Yvonna S. Lincoln. 2000. Introduction: The discipline and practice of qualitative research. Pp. 1–28 in Handbook of qualitative research, 2nd ed., eds. Norman Denzin and Yvonna S. Lincoln. Thousand

Oaks, CA: Sage.

DeParle, Jason. 1999. Project to rescue needy stumbles against the persistence of poverty. New York Times, May 15: A1, A10.

Desmond, Matthew. 2007. On the fireline: Living and dying with wildland firefighters. Chicago: University of Chicago Press.

de Vaus, David. 2008. Comparative and cross-national designs. Pp. 248–264 in The SAGE handbook of social research methods, eds. Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Diamond, Timothy. 1992. Making gray gold: Narratives of nursing home care. Chicago: University of Chicago Press.

Dill, Karen E., and Kathryn P. Thill. 2007. Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. Sex Roles 57: 851–864.

Dillman, Don A. 1978. Mail and telephone surveys: The total design method. New York: Wiley.

Dillman, Don A. 1982/1991. Mail and other self-administered questionnaires. Pp. 637–638 in Handbook of research design and social measurement, 5th ed., ed. Delbert C. Miller. Newbury Park, CA: Sage.

Dillman, Don A. 2000. Mail and Internet surveys: The tailored design method, 2nd ed. New York: Wiley.

Dillman, Don A. 2007. Mail and Internet surveys: The tailored design method, 2nd ed. Update with new Internet, visual, and mixed-mode guide. Hoboken, NJ: Wiley.

Donath, Judith S. 1999. Identity and deception in the virtual community. Pp. 29–59 in Communities in cyberspace, eds. Peter Kollock and Marc A. Smith. New York: Routledge.

Drake, Robert E., Gregory J. McHugo, Deborah R. Becker, William A. Anthony, and Robin E. Clark. 1996. The New Hampshire study of supported employment for people with severe mental illness. Journal of Consulting and Clinical Psychology 64: 391–399.

Duckworth, Kenneth, John H. Halpern, Russell K. Schutt, and Christopher Gillespie. 2003. Use of schizophrenia as a metaphor in U.S. newspapers. Psychiatric Services 54: 1402–1404.

Duneier, Mitchell. 2006. Ethnography, the ecological fallacy, and the 1995 Chicago heat wave. American Sociological Review 71(4): 679–687.

Durkheim, Émile. 1951. Suicide. New York: Free Press.

Durkheim, Émile. 1966. Suicide: A study in sociology. Translated by John A. Spaulding and George Simpson. New York: Free Press.

Durkheim, Émile. 1984. The division of labor in society. Translated by W. D. Halls. New York: Free Press.

Ekman, Paul. 2003. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. New York: Henry Holt.

Elliott, Andrew, and Daniela Nesta. 2008. Romantic red: Red enhances men's attraction to women. Journal of Personality and Social Psychology 95(5): 1150–1164.

Ellis, Carolyn. 1986. Fisher folk: Two communities on Chesapeake Bay.

Lexington: University Press of Kentucky.

Ember, Carol R., and Melvin Ember. 2011. A basic guide to cross-cultural research. New Haven, CT: Human Relations Area Files, Yale University. From http://www.yale.edu/hraf/guides.htm (accessed August 31, 2011).

Emerson, Robert M., ed. 1983. Contemporary field research. Prospect Heights, IL: Waveland Press.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. Writing ethnographic fieldnotes. Chicago: University of Chicago Press.

Erikson, Kai T. 1966. Wayward Puritans: A study in the sociology of deviance. New York: Wiley.

Erikson, Kai T. 1967. A comment on disguised observation in sociology. Social Problems 12: 366–373.

Erikson, Kai T. 1976. Everything in its path: Destruction of community in the Buffalo Creek flood. New York: Simon & Schuster.

Estabrook, Robin E., Russell K. Schutt, and Mary Lou Woodford. 2008. Translating research into practice: The participatory expert panel approach. The Open Health Services and Policy Journal 1: 19–26.

Exum, M. Lyn, Jennifer L. Hartman, Paul C. Friday, and Vivian B. Lord. 2010. Policing domestic violence in the post-SARP Era: The impact of a domestic violence police unit. Crime & Delinquency 20(10): 1–34.

Facebook. 2013. From http://newsroom.fb.com/Key-Facts (accessed December 26, 2013).

Fallon, Kathleen M., Liam Swiss, and Jocelyn Viterna. 2012. Resolving the democracy paradox: Democratization and women's legislative representation in developing nations, 1975 to 2009. American Sociological Review 77(3): 380–408.

Favreault, Melissa. 2008. Discrimination and economic mobility. The Urban Institute April 3.

Fenno Jr., Richard F. 1978. Home style: House members in their districts. Boston: Little, Brown.

Ferguson, Kristin M., Kimberly Bender, Sanna J. Thompson, Elaine M. Maccio, and David Pollio. 2012. Employment status and income generation among homeless young adults: Results from a five-city, mixed-methods study. Youth & Society 44: 385–407.

Fergusson, Ross. 2013. Risk, responsibilities and rights: Reassessing the "economic causes of crime" thesis in a recession. Youth Justice 13(1): 31–56.

File, Thom. 2013a. Computer and Internet use in the United States. Pp. 20–569 in Current Population Survey Reports. Washington, DC: U.S. Census Bureau.

File, Thom. 2013b. The diversifying electorate—Voting rates by race and Hispanic origin in 2012 (and other recent elections). Population Characteristics, Current Population Survey. Washington, DC: U.S. Census Bureau.

Fink, Arlene. 2005. Conducting research literature reviews: From the Internet to paper, 2nd ed. Thousand Oaks, CA: Sage.

*** Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker. 2011. The Oregon health insurance experiment: Evidence from the first year. National

Bureau of Economic Research Working Paper 17190.

Fowler, Floyd J. 1988. Survey research methods, rev. ed. Newbury Park, CA: Sage.

Fowler, Floyd J. 1995. Improving survey questions: Design and evaluation. Thousand Oaks, CA: Sage.

Fowler, Floyd J. 1998. Personal communication, January 7. Center for Survey Research, University of Massachusetts, Boston.

Fox, Nick, and Chris Roberts. 1999. GPs in cyberspace: The sociology of a "virtual community." The Sociological Review 47: 643–669.

Franklin, Mark N. 1996. Electoral participation. Pp. 216–235 in Comparing Democracies: Elections and Voting in Global Perspective, eds. Lawrence LeDuc, Richard G. Niemi, and Pippa Norris. Thousand Oaks, CA: Sage.

Freshman, Audrey. 2012. Financial disaster as a risk factor for posttraumatic stress disorder: Internet survey of trauma in victims of the Madoff Ponzi scheme. Health & Social Work 37: 39–48.

Freud, Sigmund. 1900/1999. The interpretation of dreams, 1st ed., ed. Ritchie Robertson, trans. Joyce Crick. Repr. Oxford, England: Oxford University Press.

Gaiser, Ted J., and Anthony E. Schreiner. 2009. A guide to conducting online research. Thousand Oaks, CA: Sage.

Gallup Organization. 2002. Poll analyses, July 29, 2002: Bush job approval update. From http://media.washingtonpost.com/wp-srv/politics/ssi/polls/postpoll_072307.html (accessed September 28, 2008).

Gallup Organization. 2008. Election polls—Accuracy record in presidential elections. From http://www.galIup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-EIections.aspx (accessed May 23, 2008).

Gallup Organization. 2011. Election Polls—Accuracy Record in Presidential Elections. From http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print (accessed March 17, 2011).

Gaubatz, K. T. (1999). Elections and war: The electoral incentive in the democratic politics of war and peace. Stanford, CA: Stanford University Press.

Gerth, Matthias A., and Gabriele Siegert. 2012. Patterns of consistence and constriction: How news media frame the coverage of direct democratic campaigns. American Behavioral Scientist 56: 279–299.

Gibson, David R. 2005. Taking turns and talking ties: Networks and conversational interaction. American Journal of Sociology 110(6): 1561–1597.

Gilbert, Dennis (with Zogby International). 2000. Hamilton College youth and guns survey. Unpublished research report.

Gilbert, Dennis (with Zogby International). 2002. Hamilton College Muslim America poll. Unpublished research report.

Gilchrist, Valerie J., and Robert L. Williams. 1999. Key informant interviews. Pp. 71–88 in Doing qualitative research, 2nd ed., eds. Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Gilligan, Carol. 1982. In a different voice: Psychological theory and women's development. Cambridge, MA: Harvard University Press.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. Nature 457 (February 19): 1012–1015.

Glaser, Barney G., and Anselm L. Strauss. 1967. The discovery of grounded theory: Strategies for qualitative research. London: Weidenfeld and Nicholson.

Glover, Judith. 1996. Epistemological and methodological considerations in secondary analysis. Pp. 28–38 in Cross-national research methods in the social sciences, eds. Linda Hantrais and Steen Mangen. New York: Pinter.

Goel, Vindu. 2014. Outcry greets Facebook's emotion test. New York Times, June 30.

Goffman, Alice. 2014. On the run: Fugitive life in an American city. Chicago: University of Chicago Press.

Goffman, Erving. 1959. Presentation of self in everyday life. New York: Doubleday Anchor.

Goffman, Erving. 1961. Asylums: Essays on the social situation of mental patients and other inmates. Garden City, NY: Doubleday.

Goffman, Erving, 1967. Interaction ritual. New York: Doubleday Anchor Books.

Goldfinger, Stephen M., and Russell K. Schutt. 1996. Comparisons of clinicians' housing recommendations of homeless mentally ill persons. Psychiatric Services 47(4): 413–415.

Goldstein, Joseph. 2014. Safer era tests "broken windows" focus on minor crime. New York Times, July 25.

Goleman, Daniel. 1993. Placebo effect is shown to be twice as powerful as expected. New York Times, August 17: C3.

Goode, Erich. 2002. Sexual involvement and social research in a fat civil rights organization. Qualitative Sociology 25(4): 501–504.

Gordon, Raymond. 1992. Basic interviewing skills. Itasca, IL: Peacock.

Griffin, Larry J. 1992. Temporality, events, and explanation in historical sociology: An introduction. Sociological Methods & Research 20: 403–427.

Griffin, Larry J. 1993. Narrative, event-structure analysis, and causal interpretation in historical sociology. American Journal of Sociology 98 (March): 1094–1133.

Groves, Robert M. 1989. Survey errors and survey costs. New York: Wiley.

Groves, Robert M., and Mick P. Couper. 1998. Nonresponse in household interview surveys. New York: Wiley.

Groves, Robert M., and Robert L. Kahn. 1979/1991. Surveys by telephone: A national comparison with personal interviews, as adapted in Delbert C. Miller's Handbook of research design and social measurement, 5th ed. Newbury Park, CA: Sage.

Gruenewald, Paul J., Andrew J. Treno, Gail Taff, and Michael Klitzner. 1997. Measuring community indicators: A systems approach to drug and alcohol problems. Thousand Oaks, CA: Sage.

Guba, Egon G., and Yvonna S. Lincoln. 1989. Fourth generation evaluation. Newbury Park, CA: Sage.

Gubrium, Jaber F., and James A. Holstein. 1997. The new language of qualitative method. New York: Oxford University Press.

Gubrium, Jaber F., and James A. Holstein. 2000. Analyzing interpretive practice. Pp. 487–508 in The handbook of qualitative research, 2nd ed., eds. Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. What the polls don't show: A closer look at U.S. church attendance. American Sociological Review 58 (December): 741–752.

Hagan, John. 1994. Crime and disrepute. Thousand Oaks, CA: Pine Forge Press.

Hage, Jerald, and Barbara Foley Meeker. 1988. Social causality. Boston: Unwin Hyman.

Hallinan, Maureen T. 1997. The sociological study of social change. American Sociological Review 62: 1–11.

Hampton, Keith N., Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. Social networking sites and our lives: How people's trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies. Washington, DC: Pew Internet & American Life Project. From http://pewinternet.org/Reports/2011/Technology-and-social-networks.aspx (accessed December 27, 2013).

Hampton, Keith N., and Neeti Gupta. 2008. Community and social interaction in the wireless city: Wi-Fi use in public and semi-public spaces. New Media & Society 10(6): 831–850.

Haney, C., C. Banks, and Philip G. Zimbardo. 1973. Interpersonal dynamics in a simulated prison. International Journal of Criminology and Penology 1:

69–97.

Hantrais, Linda, and Steen Mangen. 1996. Method and management of cross-national social research. Pp. 1–12 in Cross-national research methods in the social sciences, eds. Linda Hantrais and Steen Mangen. New York: Pinter.

Hard, Stephen F., James M. Conway, and Antonia C. Moran. 2006. Faculty and college student beliefs about the frequency of student academic misconduct. The Journal of Higher Education 77: 1058–1080.

Hart, Chris. 1998. Doing a literature review: Releasing the social science research imagination. London: Sage.

Hartnett, Kevin. 2014. Boston’s black-market cigarette problem. Boston Globe, February 2.

Heckman, James J., and Peter Siegelman. 1993. The Urban Institute Audit Studies: Their methods and findings. Pp. 187–258 in Clean and convincing evidence: Measurement of discrimination in America, eds. M. Fix and R. J. Struyk. Washington, DC: Urban Institute Press.

Herberman, Ronald B. (chair). 2005. Summary statement: Review of UMB/DFHCC Comprehensive Cancer Partnership Program. National Cancer Institute Special Emphasis Panel, June 13–June 15, National Cancer Institute.

Herek, Gregory. 1995. Developing a theoretical framework and rationale for a research proposal. Pp. 85–91 in How to write a successful research grant application: A guide for social and behavioral scientists, eds. Willo Pequegnat and Ellen Stover. New York: Plenum Press.

Hite, Shere. 1987. Women and love: A cultural revolution in progress. New York: Knopf.

Ho, D. Y. F. 1996. Filial piety and its psychological consequences. Pp. 155–165 in Handbook of Chinese psychology, ed. M. H. Bond. Hong Kong: Oxford University Press.

Hoffman, Stephen. 2014. Zero benefit: Estimating the effect of zero tolerance discipline polices on racial disparities in school discipline. Educational Policy 28(1): 69–95.

Houtzager, Peter P., and Arnab K. Acharya. 2011. Associations, active citizenship, and the quality of democracy in Brazil and Mexico. Theory and Society 40: 1–36.

Hoyle, Carolyn, and Andrew Sanders. 2000. Police response to domestic violence: From victim choice to victim empowerment. British Journal of Criminology 40: 14–26.

Huberman, A. Michael, and Matthew B. Miles. 1994. Data management and analysis methods. Pp. 428–444 in Handbook of qualitative research, eds. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Huff, Darrell. 1954. How to lie with statistics. New York: W. W. Norton.

Humphrey, Nicholas. 1992. A history of the mind: Evolution and the birth of consciousness. New York: Simon & Schuster.

Humphreys, Laud. 1970. Tearoom trade: Impersonal sex in public places. Chicago: Aldine.

Hunt, Morton. 1985. Profiles of social research: The scientific study of human interactions. New York: Russell Sage Foundation.

Huston, Patricia, and C. David Naylor. 1996. Health services research:

Reporting on studies using secondary data sources. Canadian Medical Association Journal 155: 1697–1702.

Internet World Statistics. 2012. From http://www.internetworldstats.com/stats.htm (accessed December 26, 2013.)

Inter-university Consortium for Political and Social Research. 1996. Guide to resources and services 1995–1996. Ann Arbor, MI: ICPSR.

Irvine, Leslie. 1998. Organizational ethics and field work realities: Negotiating ethical boundaries in codependents anonymous. Pp. 167–183 in Doing ethnographic research: Fieldwork settings, ed. Scott Grills. Thousand Oaks, CA: Sage.

James, Nalita, and Hugh Busher. 2009. Online interviewing. Thousand Oaks, CA: Sage.

Jankowski, Martin Sanchez. 1991. Islands in the street: Gangs and American urban society. Berkeley: University of California Press.

Jarvis, Helen. 1997. Housing, labour markets and household structure: Questioning the role of secondary data analysis in sustaining the polarization debate. Regional Studies 31: 521–531.

Jerolmack, Colin. 2007. Animal practices, ethnicity and community: The Turkish pigeon handlers of Berlin. American Sociological Review 72(6): 874–894.

Jerolmack, Colin. 2009. Primary groups and cosmopolitan ties: The rooftop pigeon flyers of New York. Ethnography 10(2/3): 211–233.

Johnson, David, and Merry Bullock. 2009. The ethics of data archiving: Issues

from four perspectives. Pp. 214–228 in The handbook of social research ethics, eds. Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Jones, James H. 1993. Bad blood: The Tuskegee syphilis experiment. New York: Free Press.

Kagay, Michael R. (with Janet Elder). 1992. Numbers are no problem for pollsters: Words are. New York Times, October 9: E5.

Kale-Lostuvali, Elif. 2007. Negotiating state provision: state-citizen encounters in the aftermath of the İzmit earthquake. The Sociological Quarterly 48: 745–767.

Kato, Yuki, Catarina Passidomo, and Daina Harvey. 2014. Political gardening in post-disaster city: Lessons from New Orleans. Urban Studies 51: 1833–1849.

Kaufman, Sharon R. 1986. The ageless self: Sources of meaning in late life. Madison: University of Wisconsin Press.

Kelling, George L., and James Q. Wilson. 1982. Broken windows: The police and neighborhood safety. Atlantic Monthly 249 (3), (March): 29–38.

Kenney, Charles. 1987. They’ve got your number. Boston Globe Magazine, August 30: 12, 46–56, 60.

Kershaw, David, and Jerilyn Fair, eds. 1976. Operations, surveys, and administration, vol. 2 of The New Jersey income-maintenance experiment. New York: Academic Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. Scientific inference in qualitative research. Princeton, NJ: Princeton University Press.

King, Nigel, and Christine Horrocks. 2010. Interviews in qualitative research. Thousand Oaks, CA: Sage.

Kinsey, Alfred C., Wardell B. Pomeroy, and Clyde E. Martin. 1948. Sexual behavior in the human male. Philadelphia: W. B. Saunders.

Kinsey, Alfred C., Wardell B. Pomeroy, Clyde E. Martin, and Paul H. Gebhard. 1953. Sexual behavior in the human female. Philadelphia: W. B. Saunders.

Kiser, Edgar, and Michael Hechter. 1991. The role of general theory in comparative-historical sociology. American Journal of Sociology 97: 1–30.

Klinenberg, Eric. 2006. Blaming the victims: Hearsay, labeling, and the hazards of quick-hit disaster ethnography. American Sociological Review 71(4): 690–698.

Klinenberg, Eric. 2012. One's a crowd. New York Times, February 5: SR4.

Klofas, John M., and Charles R. Cutshall. 1985. The social archaeology of a juvenile facility: Unobtrusive measures in the study of institutional cultures. Qualitative Sociology 8(4), (Winter): 368–387.

Koegel, Paul. 1987. Ethnographic perspectives on homeless and homeless mentally ill women. Washington, DC: Alcohol, Drug Abuse, and Mental Health Administration, Public Health Service, U.S. Department of Health and Human Services.

Kohut, Andrew. 1988. Polling: Does more information lead to better understanding? Boston Globe, November 7: 25.

Kohut, Andrew, Scott Keeter, Carroll Doherty, Michael Dimock, and Leah Christian. 2012. Assessing the representativeness of public opinion surveys.

Washington, DC: Pew Research Center. From http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/ (accessed May 25, 2014).

Kolata, Gina. 2013. Decades later, condemnation for a skid row cancer study. New York Times, October 18: A1.

Koppel, Ross. 2008. The utility of sociology. Sociological Viewpoints (Fall): 5–16.

Korn, James H. (1997). Illusions of reality: A history of deception in social psychology. Albany: State University of New York Press.

*** Kotkin, Stephen. 2002. A world war among professors. New York Times, September 7: Arts 1, 17.

Kozinets, Robert V. 2010. Netnography: Doing ethnographic research online. Thousand Oaks, CA: Sage.

Krauss, Clifford. 1996. New York crime rate plummets to levels not seen in 30 years. New York Times, December 20: A1, B4.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. Public Opinion Quarterly 72: 847–865.

Krueger, Richard A. 1988. Focus groups: A practical guide for applied research. Newbury Park, CA: Sage.

Krueger, Richard A., and Mary Anne Casey. 2000. Focus groups: A practical guide for applied research, 3rd ed. Thousand Oaks, CA: Sage.

Kuzel, Anton J. 1999. Sampling in qualitative inquiry. Pp. 33–45 in Doing qualitative research, 2nd ed., eds. Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Kvale, Steinar. 1996. Interviews: An introduction to qualitative research interviewing. Thousand Oaks, CA: Sage.

Labaw, Patricia J. 1980. Advanced questionnaire design. Cambridge, MA: ABT Books.

Labov, William. 1972. The social stratification of (r) in New York City department stores. Pp. 43–69 in Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

Lacey, John H., Tara Kelley-Baker, Robert B. Voas, Eduardo Romano, C. Debra Furr-Holden, Pedro Torres, and Amy Berning. 2011. Alcohol- and drug-involved driving in the United States: Methodology for the 2007 National Roadside Survey. Evaluation Review 35: 319–353.

Lange, Matthew. 2013. Comparative–historical methods. Thousand Oaks, CA: Sage.

LaRossa, Ralph. Parenthood in early twentieth-century America project (PETCAP), 1900–1944 [Computer file]. Atlanta: Georgia State University [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1997.

Larson, Calvin J. 1993. Pure and applied sociological theory: Problems and issues. New York: Harcourt Brace Jovanovich.

Latour, Francie. 2002. Marching orders: After 10 years, state closes prison boot camp. Boston Sunday Globe, June 16: Bl, B7.

Lavin, Michael R. 1994. Understanding the 1990 Census: A guide for marketers, planners, grant writers and other data users. Kenmore, NY: Epoch Books.

Lavrakas, Paul J. 1987. Telephone survey methods: Sampling, selection, and supervision. Newbury Park, CA: Sage.

LeDuc, Lawrence, Richard G. Niemi, and Pippa Norris (Eds.). 1996. Comparing democracies: Elections and voting in global perspective. Thousand Oaks, CA: Sage.

Legal Information Institute. 2006. U.S. Code Collection: Title 42, Chapter 7, Subchapter XI, Part C, §I320d. From the Cornell University Law School website: http://www4.1aw.cornell.edu/uscode/html/uscode42/usc_sec_42_0000I320—d000-.html (accessed September 28, 2008).

Lelieveldt, Herman. 2003. March–April. Increasing social capital through direct democracy? A case study of the "It's Our Neighbourhood's Turn" project. Paper presented at the ECPR-Joint Sessions (Workshop 22: Bringing citizens back in: participatory democracy and political participation), Edinburgh, Scotland. From http://www.paltin.ro/biblioteca/Lelieveldt.pdf (accessed September 28, 2008).

Lelieveldt, Herman. 2004. Helping citizens help themselves: Neighborhood improvement programs and the impact of social networks, trust, and norms on neighborhood-oriented forms of participation. Urban Affairs Review 39: 531–551.

Lempert, Richard. 1989. Humility is a virtue: On the publicization of policy-relevant research. Law & Society Review 23(1): 145–161.

Lempert, Richard, and Joseph Sanders. 1986. An invitation to law and social science: Desert, disputes, and distribution. New York: Longman.

Levy, Paul S., and Stanley Lemeshow. 1999. Sampling of populations: Methods and applications, 3rd ed. New York: Wiley.

Lewin, Tamar. 2001a. Surprising result in welfare-to-work studies. New York Times, July 31: A16.

Lewin, Tamar. 2001b. Income education is found to lower risk of new arrest. New York Times, November 16: A18.

Lieberson, Stanley. 1985. Making it count: The improvement of social research and theory. Berkeley: University of California Press.

Ling, Rich, and Gitte Stald. 2010. Mobile communities: Are we talking about a village, a clan, or a small group? American Behavioral Scientist 53(8): 1113–1147.

Locke, Lawrence F., Stephen J. Silverman, and Waneen Wyrick Spirduso. 1998. Reading and understanding research. Thousand Oaks, CA: Sage.

Locke, Lawrence F., Waneen Wyrick Spirduso, and Stephen J. Silverman. 2000. Proposals that work: A guide for planning dissertations and grant proposals, 4th ed. Thousand Oaks, CA: Sage.

Lockwood, Daniel. Violent incidents among selected public school students in two large cities of the South and the Southern Midwest, 1995 [United States] [Computer file]. ICPSR version. Atlanta: Clark Atlantic University [producer], 1996. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1998.

Luker, Kristin. 1985. Abortion and the politics of motherhood. Berkeley: University of California Press.

Makarios, Matthew D., and Travis C. Pratt. 2012. The effectiveness of policies and programs that attempt to reduce firearm violence: A meta-analysis. Crime & Delinquency 58(2): 222–244.

Mangione, Thomas W. 1995. Mail surveys: Improving the quality. Thousand Oaks, CA: Sage.

Margolis, Eric. 2004. Looking at discipline, looking at labour: Photographic representations of Indian boarding schools. Visual Studies 19: 72–96.

Marini, Margaret Mooney, and Burton Singer. 1988. Causality in the social sciences. Pp. 347–409 in Sociological methodology, vol. 18, ed. Clifford C. Clogg. Washington, DC: American Sociological Association.

Markoff, John. 2005. Transitions to democracy. Pp. 384–403 in The handbook of political sociology: States, civil societies, and globalization, eds. Thomas Janoski, Robert R. Alford, Alexander M. Hicks, and Mildred A. Schwartz. New York: Cambridge University Press.

Marsden, Peter V. 1987. Core discussion networks of Americans. American Sociological Review 52: 122–131.

Marshall, S. L. A. 1947/1978. Men against fire, reprint ed. Gloucester, MA: Peter Smith.

Martin, Lawrence L., and Peter M. Kettner. 1996. Measuring the performance of human service programs. Thousand Oaks, CA: Sage.

Martin, Linda G., and Kevin Kinsella. 1995. Research on the demography of aging in developing countries. Pp. 356–403 in Demography of aging, eds. Linda G. Martin and Samuel H. Preston. Washington, DC: National Academy Press.

Marx, Karl. 1967. Capital: A critique of political economy. New York: International Publishers.

Maxwell, Joseph A. 1996. Qualitative research design: An interactive approach. Thousand Oaks, CA: Sage.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. Big Data: A revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt.

McCarter, Susan A. 2009. Legal and extralegal factors affecting minority overrepresentation in Virginia's juvenile justice system: A mixed-methods study. Child and Adolescent Social Work Journal 26: 533–544.

McLellan, A. Thomas, Lester Luborsky, John Cacciola, Jeffrey Griffith, Frederick Evans, Harriet L. Barr, and Charles P. O'Brien. 1985. New data from the Addiction Severity Index: Reliability and validity in three centers. The Journal of Nervous and Mental Disease 173(7): 412–423.

McPherson, Miller, Lynn Smith-Lovin, and Matthew E. Brashears. 2006. Social isolation in America: Changes in core discussion networks over two decades. American Sociological Review 71: 353–375.

Mead, Margaret. 1928/2001. Coming of age in Samoa, reprint ed. New York: HarperCollins Perennial.

Melbin, Murray. 1978. Night as frontier. American Sociological Review 43(1): 3–22.

Milbrath, Lester, and M. L. Goel. 1977. Political participation, 2nd ed. Chicago: Rand McNally.

Miles, Matthew B., and A. Michael Huberman. 1994. Qualitative data analysis, 2nd ed. Thousand Oaks, CA: Sage.

Milgram, Stanley. 1963. Behavioral study of obedience. Journal of Abnormal and Social Psychology 67: 371–478.

Milgram, Stanley. 1964. Issues in the study of obedience: A reply to Baumrind. American Psychologist 19: 848–852.

Milgram, Stanley. 1965. Some conditions of obedience and disobedience to authority. Human Relations 18: 57–76.

Milgram, Stanley. 1974. Obedience to authority: An experimental view. New York: Harper & Row.

Miller, Arthur G. 1986. The obedience experiments: A case study of controversy in social science. New York: Praeger.

Miller, Delbert C. 1991. Handbook of research design and social measurement, 5th ed. Newbury Park, CA: Sage.

Miller, Delbert C., and Nell J. Salkind. 2002. Handbook of research design and social measurement, 6th ed. Thousand Oaks, CA: Sage.

Miller, Susan. 1999. Gender and community policing: Walking the talk. Boston: Northeastern University Press.

Miller, William L., and Benjamin F. Crabtree. 1999. The dance of interpretation. Pp. 127–143 in Doing qualitative research, eds. Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Mitchell Jr., Richard G. 1993. Secrecy and fieldwork. Newbury Park, CA: Sage.

Mohr, Lawrence B. 1992. Impact analysis for program evaluation. Newbury Park, CA: Sage.

Monkkonen, Eric H. 1994. Introduction. Pp. 1–8 in Engaging the past: The uses of history across the social sciences. Durham, NC: Duke University Press.

Mooney, Christopher Z., and Mei Hsien Lee. 1995. Legislating morality in the American states: The case of abortion regulation reform. American Journal of Political Science 39: 599–627.

Moore Jr., Barrington. 1966. Social origins of democracy and dictatorship: Lord and peasant in the making of the modern world. Boston: Beacon Press.

Moore, Spencer, Mark Daniel, Laura Linnan, Marci Campbell, Salli Benedict, and Andrea Meier. 2004. After Hurricane Floyd passed: Investigating the social determinants of disaster preparedness and recovery. Family and Community Health 27: 204–217.

Morrill, Calvin, Christine Yalda, Madeleine Adelman, Michael Musheno, and Cindy Bejarano. 2000. Telling tales in school: Youth culture and conflict narratives. Law & Society Review 34: 521–565.

Mullins, Carolyn J. 1977. A guide to writing and publishing in the social and behavioral sciences. New York: Wiley.

Myrdal, Gunnar. 1944/1964. An American dilemma, reprint ed. New York: McGraw-Hill.

National Geographic Society. 2000. Survey 2000. From http://business.clemson.edu/socio/s2k data211.html (accessed September 28,

2008).

National Opinion Research Center (NORC). 1992. National data program for the social sciences: The NORC General Social Survey; Questions and answers (mimeographed). Chicago: NORC.

National Opinion Research Center (NORC). 2006. General social survey. Chicago: National Opinion Research Center, University of Chicago.

National Opinion Research Center (NORC). 2014. General social survey. From http://www3.norc.org/GSS+Website/Publications/GSS+Questionnaires/ (accessed May 25, 2014).

Navarro, Mireya. 2012. For many Latinos, racial identity is more culture than color. New York Times, January 14: A9.

Needleman, Carolyn. 1981. Discrepant assumptions in empirical research: The case of juvenile court screening. Social Problems 28 (February): 247–262.

Neuendorf, Kimberly A. 2002. The content analysis guidebook. Thousand Oaks, CA: Sage.

Neumark, David. 1996. Sex discrimination in restaurant hiring: An audit study. Quarterly Journal of Economics 111: 915–941.

Newbury, Darren. 2005. Editorial: The challenge of visual studies. Visual Studies 20: 1–3.

Newman, Katherine S., Cybelle Fox, David J. Harding, Jal Mehta, and Wendy Roth. 2004. Rampage: The social roots of school shootings. New York: Basic Books.

O'Dochartaigh, Niall. 2002. The Internet research handbook: A practical guide for students and researchers in the social sciences. Thousand Oaks, CA: Sage.

Office of Management and Budget. (n.d.). Government Performance Results Act of 1993. From www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html (accessed September 28, 2008).

Olzak, Susan, Suzanne Shanahan, and Elizabeth H. McEneaney. 1996. Poverty, segregation, and race riots: 1960 to 1993. American Sociological Review 61: 590–613.

Onoye, Jane M., Deborah A. Goebert, and Stephanie T. Nishimura. 2012. Use of incentives and Web-based administration for surveying student alcohol and substance use in an ethnically diverse sample. Journal of Substance Use 17(1): 61–71.

Orcutt, James D., and J. Blake Turner. 1993. Shocking numbers and graphic accounts: Quantified images of drug problems in the print media. Social Problems 49: 190–206.

Orr, Larry L. 1999. Social experiments: Evaluating public programs with experimental methods. Thousand Oaks, CA: Sage.

Pager, Devah. 2003. The mark of a criminal record. American Journal of Sociology 108: 937–975.

Pager, Devah, and Lincoln Quillian. 2005. Walking the talk? What employers say versus what they do. American Sociological Review 70: 355–380.

Pagnini, Deanna L., and S. Philip Morgan. 1996. Racial differences in marriage and childbearing: Oral history evidence from the South in the early twentieth century. American Journal of Sociology 101: 1694–1715.

Papineau, David. 1978. For science in the social sciences. London: Macmillan.

Parker-Pope, Tara. 2008. Love, sex, and the changing landscape of infidelity. New York Times, October 28: D1.

Parks, Kathleen A., Ann M. Pardi, and Clara M. Bradizza. 2006. Collecting data on alcohol use and alcohol-related victimization: A comparison of telephone and web-based survey methods. Journal of Studies on Alcohol 67: 318–323.

Parks, Malcolm R., and Kory Floyd. 1996. Making friends in cyberspace. Journal of Communication 46(1): 80–97.

Parlett, Malcolm, and David Hamilton. 1976. Evaluation as illumination: A new approach to the study of innovative programmes. Pp. 140–157 in Evaluation studies review annual, vol. 1, ed. G. Class. Beverly Hills, CA: Sage.

Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. Do fair procedures matter? The effect of procedural justice on spouse assault. Law & Society Review 31(1): 163–204.

Patton, Michael Quinn. 2002. Qualitative research & evaluation methods, 3rd ed. Thousand Oaks, CA: Sage.

Paxton, Pamela. 2002. Social capital and democracy: An interdependent relationship. American Sociological Review 67: 254–277.

Perry, Gina. 2013. Behind the shock machine: The untold story of the notorious Milgram psychology experiments. New York: New Press.

Pew Internet & American Life Project. 2010. Change in Internet access by age group, 2000–2009. Washington, DC: Pew Internet & American Life Project. From http://www.pewinternet.org/Infographics/2010/Internet-acess-by-age-

group-over-time.aspx# (accessed May 15, 2011).

Pew Research Center. 2013. Social networking use. From http://www.pewresearch.org/data-trend/media-and-technology/social-networking-use/ (accessed December 26, 2013).

Phillips, David P. 1982. The impact of fictional television stories on U.S. adult fatalities: New evidence on the effect of the mass media on violence. American Journal of Sociology 87 (May): 1340–1359.

Phoenix, Ann. 2003. Neoliberalism and masculinity: Racialization and the contradictions of schooling for 11- to 14-year-olds. Youth & Society 36: 227–246.

Piliavin, Jane Allyn, and Irving M. Piliavin. 1972. Effect of blood on reactions to a victim. Journal of Personality and Social Psychology 23: 353–361.

Pipher, Mary. 1994. Reviving Ophelia: Saving the selves of adolescent girls. New York: Ballantine Books.

Pollan, Michael. 2008. In defense of food: An eater's manifesto. New York: Penguin Press.

Posavac, Emil J., and Raymond G. Carey. 1997. Program evaluation: Methods and case studies, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Presley, Cheryl A., Philip W. Meilman, and Rob Lyerla. 1994. Development of the core alcohol and drug survey: Initial findings and future directions. Journal of American College Health 42: 248–255.

Price, Richard H., Michelle Van Ryn, and Amiram D. Vinokur. 1992. Impact of a preventive job search intervention on the likelihood of depression among the

unemployed. Journal of Health and Social Behavior 33 (June): 158–167.

Punch, Maurice. 1994. Politics and ethics in qualitative research. Pp. 83–97 in Handbook of qualitative research, eds. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Putnam, Robert D. 2000. Bowling alone: The collapse and revival of American community. New York: Touchstone.

Pyrczak, Fred. 2005. Evaluating research in academic journals: A practical guide to realistic evaluation, 3rd ed. Glendale, CA: Pyrczak.

Radin, Charles A. 1997. Partnerships, awareness behind Boston's success. Boston Globe, February 19: A2, B7.

Ragin, Charles C. 1987. The comparative method: Moving beyond qualitative and quantitative strategies. Berkeley: University of California Press.

Ragin, Charles C. 1994. Constructing social research. Thousand Oaks, CA: Pine Forge Press.

Ragin, Charles C. 2000. Fuzzy-set social science. Chicago: University of Chicago Press.

Ramirez, Anthony. 2002. One more reason you're less likely to be murdered. New York Times, August 25: WK3.

Rankin, Bruce H., and James M. Quane. 2002. Social contexts and urban adolescent outcomes: The interrelated effects of neighborhoods, families, and peers on African-American youth. Social Problems 49: 79–100.

Regnerus, Mark. 2007. Forbidden fruit: Sex and religion in the lives of American teenagers. New York: Oxford University Press.

Reiss Jr., Albert J. 1971. The police and the public. New Haven, CT: Yale University Press.

Rew, Lynn, Deborah Koniak-Griffin, Mary Ann Lewis, Margaret Miles, and Ann O'Sullivan. 2000. Secondary data analysis: New perspective for adolescent research. Nursing Outlook 48: 223–229.

Reynolds, Paul Davidson. 1979. Ethical dilemmas and social science research San Francisco: Jossey-Bass.

Richards, Thomas J., and Lyn Richards. 1994. Using computers in qualitative research. Pp. 445–462 in Handbook of qualitative research, eds. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Richardson, Laurel. 1995. Narrative and sociology. Pp. 198–221 in Representation in ethnography, ed. John Van Maanen. Thousand Oaks, CA: Sage.

Riedel, Marc. 2000. Research strategies for secondary data: A perspective for criminology and criminal justice. Thousand Oaks, CA: Sage.

Riessman, Catherine Kohler. 2002. Narrative analysis. Pp. 217–270 in The qualitative researcher's companion, eds. A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage.

Rinehart, Jenny K., and Elizabeth A. Yeater. 2011. A qualitative analysis of sexual victimization narratives. Violence Against Women 17(7): 925–943.

Ringwalt, Christopher L., Jody M. Greene, Susan T. Ennett, Ronaldo Iachan,

Richard R. Clayton, and Carl G. Leukefeld. 1994. Past and future directions of the D.A.R.E. program: An evaluation review. Research Triangle Park, NC: Research Triangle Institute.

Rives Jr., Norfleet W., and William J. Serow. 1988. Introduction to applied demography: Data sources and estimation techniques. SAGE University Paper Series on Quantitative Applications in the Social Sciences, series No. 07–039. Thousand Oaks, CA: Sage.

Robertson, David Brian. 1993. The return to history and the new institutionalism in American political science. Social Science History 17: 1–36.

Roman, Anthony. 2005. Women's health network client survey: Field report. Unpublished report. Boston: Center for Survey Research, University of Massachusetts.

Rosen, Lawrence. 1995. The creation of the Uniform Crime Report: The role of social science. Social Science History 19: 215–238.

Rosenberg, Morris. 1968. The logic of survey analysis. New York: Basic Books.

Rosenbloom, Stephanie. 2007. On Facebook, scholars link up with data. New York Times, December 17.

Rossi, Peter H., and Howard E. Freeman. 1989. Evaluation: A systematic approach, 4th ed. Newbury Park, CA: Sage.

Rossman, Gretchen B., and Sharon F. Rallis. 1998. Learning in the field: An introduction to qualitative research. Thousand Oaks, CA: Sage.

Rubin, Herbert J., and Irene S. Rubin. 1995. Qualitative interviewing: The art of hearing data. Thousand Oaks, CA: Sage.

Ruderman, Wendy. 2012. Crime report manipulation is common among New York police, study finds. New York Times, June 29: A17.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. Capitalist development and democracy. Chicago: University of Chicago Press.

Sacks, Stanley, Karen McKendrick, George DeLeon, Michael T. French, and Kathryn E. McCollister. 2002. Benefit-cost analysis of a modified therapeutic community for mentally ill chemical abusers. Evaluation & Program Planning 25: 137–148.

Salisbury, Robert H. 1975. Research on political participation. American Journal of Political Science 19 (May): 323–341.

Sampson, Robert J., and Janet L. Lauritsen. 1994. Violent victimization and offending: individual-, situational-, and community-level risk factors. Pp. 1–114 in Social influences, vol. 3 of Understanding and preventing violence, eds. Albert J. Reiss Jr. and Jeffrey A. Roth. Washington, DC: National Academies Press.

Sampson, Robert J., and John H. Laub. 1993. Structural variations in juvenile court processing: Inequality, the underclass, and social control. Law & Society Review 27(2): 285–311.

Sampson, Robert J., and John H. Laub. 1994. Urban poverty and the family context of delinquency: A new look at structure and process in a classic study. Child Development 65: 523–540.

Sampson, Robert J., and Stephen W. Raudenbush. 1999. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. American Journal of Sociology 105: 603–651.

Savage, Charlie. 2011. Trend to lighten harsh sentences catches on in

conservative states. New York Times, August 13: A12.

Scarce, Rik. 2005. Eco-warriors: Understanding the radical environmental movement, 2nd ed. Walnut Creek, CA: Left Coast Press.

Schaeffer, Nora Cate, and Stanley Presser. 2003. The science of asking questions. Annual Review of Sociology 29: 65–88.

Schalock, Robert, and John Butterworth. 2000. A benefit-cost analysis model for social service agencies. Boston: Institute for Community Inclusion.

Schapira, Lidia, and Russell K. Schutt. 2011. Training community health workers about cancer clinical trials. Journal of Immigrant and Minority Health. From http://www.springerlink.com/content/g6p05g3113951q11/. (accessed November 22, 2014).

Schober, Michael F. 1999. Making sense of survey questions. Pp. 77–94 in Cognition and survey research, eds. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwartz, Judith M. Tanur, and Roger Tourangeau. New York: Wiley.

Schorr, Lisbeth B., and Daniel Yankelovich. 2000. In search of a gold standard for social programs. Boston Globe, February 18: A14.

Schuck, Amie M. 2013. A life-course perspective on adolescents' attitudes to police: DARE, delinquency, and residential segregation. Journal of Research in Crime and Delinquency 50(4): 579–607.

Schuman, Howard, and Stanley Presser. 1981. Questions and answers in attitude surveys: Experiments on question form, wording, and context. New York: Academic Press.

Schutt, Russell K. 1986. Organization in a changing environment. Albany: State University of New York Press.

Schutt, Russell K. 1988. Working with the homeless: The backgrounds, activities and beliefs of shelter staff. Boston: University of Massachusetts. Unpublished report.

Schutt, Russell K. (with Stephen M. Goldfinger) 2011. Homelessness, housing and mental illness. Cambridge, MA: Harvard University Press.

Schutt, Russell K., Elizabeth Riley Cruz, and Mary Lou Woodford. 2008. Client satisfaction in a breast and cervical cancer early detection program: The influence of ethnicity and language, health, resources, and barriers. Women & Health 48: 283–302.

Schutt, Russell K., and Jacqueline Fawcett. 2005. Case management in the women's health network. Boston: University of Massachusetts. Unpublished report.

Schutt, Russell K., Jacqueline Fawcett, Gail B. Gall, Brooke Harrow, and Mary Lou Woodford. 2010. Case manager satisfaction in public health. Professional Case Management 15: 124–134.

Schutt, Russell K., Jessica Santiccioli, Jennifer Maniates, Silas Henlon, Lidia Schapira. 2008. Community health workers and cancer clinical trials: Experiences, knowledge, and concerns. Research report for project funded by National Cancer Institute.

Schutt, Russell, JudyAnn Bigby, and Lidia Schapira. 2005. Educating underserved communities about cancer clinical trials. Proposal to the National Cancer Institute. Boston: University of Massachusetts Boston and Dana-Farber/Harvard Cancer Center.

Schutt, Russell K., Lidia Schapira, Jennifer Maniates, Jessica Santiccioli, Silas Henlon, JudyAnn Bigby. 2010. Community health workers' support for cancer clinical trials: Description and explanation. Journal of Community Health 35: 417–422.

Schutt, Russell K., and M. L. Fennell. 1992. Shelter staff satisfaction with services, the service network, and their jobs. Current Research on Occupations and Professions 7: 177–200.

Schutt, Russell K., and Stephen M. Goldfinger. 1996. Housing preferences and perceptions of health and functioning among mentally ill persons. Psychiatric Services 47(4): 381–386.

Schutt, Russell K., Xiaogang Deng, Gerald R. Garrett, Stephanie Hartwell, Sylvia Mignon, Joseph Bebo, Matthew O'Neill, Mary Aruda, Pat Duynstee, Pam DiNapoli, and Helen Reiskin. 1996. Substance use and abuse among UMass Boston students. Unpublished report, Department of Sociology, University of Massachusetts, Boston.

Scriven, Michael. 1972. Prose and cons about goal-free evaluation. Evaluation Comment 3: 1–7.

Selm, Martine Van, and Nicholas W. Jankowski. 2006. Conducting online surveys. Quality & Quantity 40: 435–456.

Shepherd, Jane, David Hill, Joel Bristor, and Pat Montalvan. 1996. Converting an ongoing health study to CAPI: Findings from the national health and nutrition study. Pp. 159–164 in Health Survey Research Methods Conference Proceedings, ed. Richard B. Warnecke. Hyattsville, MD: U.S. Department of Health and Human Services.

Sherman, Lawrence W. 1992. Policing domestic violence: Experiments and dilemmas. New York: Free Press.

Sherman, Lawrence W. 1993. Implications of a failure to read the literature. American Sociological Review 58: 888–889.

Sherman, Lawrence W., and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. American Sociological Review 49: 261–272.

Sieber, Joan E. 1992. Planning ethically responsible research: A guide for students and internal review boards. Thousand Oaks, CA: Sage.

Singer, Eleanor. 2006. Introduction: Nonresponse bias in household surveys. Public Opinion Quarterly 70(5): 637–645.

Sjoberg, Gideon, ed. 1967. Ethics, politics, and social research. Cambridge, MA: Schenkman.

Skocpol, Theda. 1984. Emerging agendas and recurrent strategies in historical sociology. Pp. 356–391 in Vision and method in historical sociology, ed. Theda Skocpol. New York: Cambridge University Press.

Smith, Tom W. 1984. Nonattitudes: A review and evaluation. Pp. 215–255 in Surveying subjective phenomena, vol. 2, eds. Charles F. Turner and Elizabeth Martin. New York: Russell Sage Foundation.

Smith, Tom. 1987. That which we call welfare by any other name would smell sweeter: An analysis of the impact of question wording on response patterns. Public Opinion Quarterly 51(1): 75–83.

St. Jean, Peter K. B. 2007. Pockets of crime: Broken windows, collective efficacy, and the criminal point of view. Chicago: University of Chicago Press.

Stake, Robert E., ed. 1975. Evaluating the arts in education: A responsible approach. Columbus, OH: Merrill.

Stake, Robert E. 1995. The art of case study research. Thousand Oaks, CA: Sage.

Statistic Brain. 2013. Social networking statistics. From http://www.statisticbrain.com/social-networking-statistics/ (accessed December 27, 2013).

Stille, Alexander. 2000. A happiness index with a long reach: Beyond G.N.P. to subtler measures. New York Times, May 20: A17, A19.

Straus, Murray, and Richard Gelles. 1988. Intimate violence. New York: Simon & Schuster.

Strickling, Lawrence E. 2010. Digital nation: 21st century America's progress toward universal broadband Internet access. An NTIA research preview. Washington, DC: National Telecommunications and Information Administration, U.S. Department of Commerce.

Strunk Jr., William, and E. B. White. 1979. The elements of style, 3rd ed. New York: Macmillan.

Sudman, Seymour. 1976. Applied sampling. New York: Academic Press.

Sue, Valerie M., and Lois A. Ritter. 2012. Conducting online surveys, 2nd ed. Thousand Oaks, CA: Sage.

Survey.net. 2000. Year 2000 presidential election survey. From http://www.survey.net/sv-press.htm. (accessed November 22, 2014).

Survey on adultery: “I do” means “I don’t.” 1993. New York Times, October 19: A20.

Tavernise, Sabrina. 2011. Youth, mobility, and poverty help drive cellphone-only status. New York Times, April 21: A13.

Taylor, Jerry. 1999. DARE gets updated in some area schools, others drop program. Boston Sunday Globe, May 16: 1, 11.

Testa, Maria, Jennifer A. Livingston, and Carol VanZile-Tamsen. 2011. Advancing the study of violence against women using mixed methods: Integrating qualitative methods into a quantitative research program. Violence Against Women 17: 236–250.

Thompson, Clive. 2007. Community urinalysis. New York Times Magazine, December 9. From http://www.nytimes.com/2007/12/09/magazine/09_11_urinalysis.html?ref=magazine (accessed September 28, 2008).

Thorne, Barrie. 1993. Gender play: Girls and boys in school. New Brunswick, NJ: Rutgers University Press.

Tierney, John. 2009. Public policy that makes test subjects of us all. New York Times, April 7: D1.

Tjaden, Patricia, and Nancy Thoennes. 2000. Extent, nature, and consequences of intimate partner violence: Findings from the National Violence Against Women Survey (NCJ 181867). Washington, DC: Office of Justice Programs, National Institute of Justice and the Centers for Disease Control and Prevention.

Toppo, Greg. 2002. Antidrug program backed by study. Boston Globe, October 29: A10.

Tourangeau, Roger. 2004. Survey research and societal change. Annual review of psychology 55: 775–801.

Tourangeau, Roger, Frederick G. Conrad, and Mick P. Couper. 2012. The science of web surveys. Oxford: Oxford University Press.

Townsend, Meg, Dana Hunt, William Rhodes. 2005. Interim report: Evaluability assessment of the President's Family Justice Center initiative. Cambridge, MA: Abt Associates. From www.ncjrs.gov/pdffiles1/nij/grants/212278.pdf (accessed March 20, 2011).

Tufte, Edward R. 1983. The visual display of quantitative information. Cheshire, CT: Graphics Press.

Turabian, Kate L. 1967. A manual for writers of term papers, theses, and dissertations, 3rd rev. ed. Chicago: University of Chicago Press.

Turner, Charles F., and Elizabeth Martin, eds. 1984. Surveying subjective phenomena. 2 vols. New York: Russell Sage Foundation.

U.S. Bureau of Economic Analysis, Communications Division. 2004. Customer satisfaction survey report, FY 2004. Washington, DC: U.S. Department of Commerce. From http://www.bea.gov/bea/about/cssr_2004_complete.pdf (accessed September 28, 2008).

U.S. Bureau of Labor Statistics, Department of Labor. 1991. Major programs of the bureau of labor statistics. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997a. Employment and earnings. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997b. Handbook of methods. Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Census Bureau. 1999. United States census 2000, updated summary: Census 2000 Operational Plan. Washington, DC: U.S. Department of Commerce, Bureau of the Census, February.

U. S. Census Bureau. 2004–2005. Statistical abstract of the United States. Washington, DC: U.S. Department of Commerce, Bureau of the Census. From http://www.census.gov/prod/www/statistical-abstract-2001_2005.html (accessed September 28, 2008).

U.S. Census Bureau. 2010. Internet use in the United States: October 2009. Washington, DC: U.S. Department of Commerce, Bureau of the Census. From http://www.census.gov/population/www/socdemo/computer/2009.html (accessed March 18, 2011).

U.S. Census Bureau. 2013. U.S. and World Population Clock. From http://www.census.gov/popclock/ (accessed December 27, 2013).

U.S. Department of Health, Education, and Welfare. 1979. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. Washington, DC: Superintendent of Documents, U.S. Government Printing Office. (ERIC Document Reproduction Service No. ED 183582.)

U.S. Government Accounting Office. 2001. Health and human services: Status of achieving key outcomes and addressing major management challenges (GAO -01–748). From http://www.gao.gov/new.items/d01748.pdf (accessed September 28, 2008).

van de Vijver, Fons, and Kwok Leung. 1997. Methods and data analysis for cross-cultural research. Thousand Oaks, CA: Sage.

Van Maanen, John. 1995. An end to innocence: The ethnography of ethnography. Pp. 1–35 in Representation in ethnography, ed. John Van Maanen. Thousand Oaks, CA: Sage.

Verba, Sidney, and Norman Nie. 1972. Political participation: Political democracy and social equality. New York: Harper & Row.

Vernez, Georges M., Audrey Burnam, Elizabeth A. McGlynn, Sally Trude, and Brian S. Mirttman. 1988. Review of California's program for the homeless mentally disabled (R-3631-CDMH). Santa Monica, CA: RAND.

Vincus, Amy A., Chris Ringwalt, Melissa S. Harris, and Stephen R. Shamblen. 2010. A short-term, quasi-experimental evaluation of D.A.R.E.'s revised elementary school curriculum. Journal of Drug Education 40: 37–49.

Wageman, Ruth. 1995. Interdependence and group effectiveness. Administrative Science Quarterly 40: 145–180.

Wallgren, Anders, Britt Wallgren, Rolf Persson, Ulf Jorner, and Jan-Aage Haaland. 1996. Graphing statistics and data: Creating better charts. Thousand Oaks, CA: Sage.

Walters, Pamela Barnhouse, David R. James, and Holly J. McCammon. 1997. Citizenship and public schools: Accounting for racial inequality in education for the pre- and post-disfranchisement South. American Sociological Review 62: 34–52.

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 2000. Unobtrusive measures, rev. ed. Thousand Oaks, CA: Sage.

Weber, Max. 1930/1992. The Protestant ethic and the spirit of capitalism, trans. Talcott Parsons. Repr. London: Routledge.

Weber, Max. 1947/1997. The theory of social and economic organization, trans. A. M. Henderson and Talcott Parsons, reprint ed. New York: Free Press.

Weber, Robert Philip. 1990. Basic content analysis, 2nd ed. Newbury Park, CA: Sage.

Weintraub, Karen. 2013. Monitoring social media to cut the military suicide rate. New York Times, July 22: B5.

Weitzman, Eben, and M. B. Miles. 1994. Computer programs for qualitative data analysis. Thousand Oaks, CA: Sage.

Wells, William, Michael R. Cavanaugh, Jeffrey A. Bouffard, and Matt R. Nobles. 2012. Non-response bias with a web-based survey of college students: Differences from a classroom survey about carrying concealed handguns. Journal of Quantitative Criminology 28: 455–476.

West, Steven L., and Ken K. O'Neal. 2004. Project D.A.R.E. outcome effectiveness revisited. American Journal of Public Health 94: 1027–1029.

Wholey, J. S. 1987. Evaluability assessment: Developing program theory. Pp. 77–92 in Using program theory in evaluation: New directions for program evaluation (no. 33), ed. Leonard Bickman. San Francisco: Jossey-Bass.

Whyte, William Foote. 1955. Street corner society. Chicago: University of Chicago Press.

Witte, James C., Lisa M. Amoroso, and Philip E. N. Howard. 2000. Research methodology: Method and representation in internet-based survey tools—mobility, community, and cultural identity in survey 2000. Social Science Computer Review 18: 179–195.

Wolcott, Harry F. 1995. The art of fieldwork. Walnut Creek, CA: AltaMira Press.

Yardley, William. 2008. Drawing lots for health care. New York Times, March 13. From http://www.nytimes.com/2008/03/13/us/13bend.html (accessed September 28, 2008).

Zaret, David. 1996. Petitions and the “invention” of public opinion in the English Revolution. American Journal of Sociology 101: 1497–1555.

Zimbardo, Philip. 2007. The Lucifer effect: Understanding how good people turn evil. New York: Random House.

Index

Abortion and the Politics of Motherhood (Luker), 222
Abt Associates, 301
Academic OneFile, 346–347
Acharya, Arnab, 276
Action research, 295
Active listening, 217
Addiction Severity Index (ASI), 82
Adler, Patricia A., 8–9, 105
Adler, Peter, 8–9
Adolescent outcomes, influence of context on, 10
Adorno, Theodor W., 82
Aggregate matching, 121
Aging, Status, and Sense of Control Survey (ASOC), 154 (box)
AIDS, 49–50, 58
Alameda Family Justice Center, 301, 302 (exhibit)–303 (exhibit)
Alcohol treatment program, 294 (exhibit)
Alfred, Randall, 207
Alternate-forms reliability, 83
Ambwani, Suman, 265, 268, 269
American Civil Liberties Union (ACLU), 299
American Dilemma, An (Myrdal), 21
American Evaluation Association, 291
American Political Science Association, 47
American Psychological Association, 47
American Sociological Association (ASA), 34
    Code of Ethics, 47–48, 131–132, 334
Anderson, Elijah, 202–203, 205, 231, 235, 240–241
Anomalous, 27
Anonymity, 159–160
Anspach, Renee, 249
Applied reports, 331, 332 (exhibit)–333 (exhibit)
Appreciative inquiry, 295
Archives, 261–263 (exhibit)
Aronson, Elliot, 56, 131

Aseltine, Robert H., Jr.
Association, 112–113
    direction of, 26
    measure of, 184–185
Asynchronous online interviewing, 219–220
Audit (paired testing), 129
Authoritarian Personality, The (Adorno et al.), 82
Availability (convenience) sampling, 103–104, 156
Aykanian, Amanda, 203 (box)

Back matter (reports), 331
Bail, Christopher, 277–278
Bar chart, 171 (exhibit)
Base number (*N*), 174, 183
Baumrind, Diana, 49, 51, 58, 59
Becker, Howard, 103, 235, 330
Before-and-after design, 120, 121 (exhibit), 122 (exhibit)–123
Bellah, Robert, 217–218
Belmont Report, 47 (exhibit)
Bender, Kimberly, 95 (box)
Bendix, Reinhard, 272, 282
Beneficence, 47 (exhibit)
Berk, Richard A., 26, 60–61, 112, 129, 132, 297
Berning, Amy, 57 (box)
Bertamini, Macro, 329
Beyond Caring (Chambliss), 75, 200–201
Bias
    archival data and, 262
    coverage, in web surveys, 155–156
    sampling, 96
    selection, 125
    survey questions and, 141–142
Bierman, Alex, 154 (b0x)
Bigby, JudyAnn, 322–328
Big Data, 190–192
Bimodal distribution, 177
Binge drinking, 69, 70 (exhibit), 333 (exhibit)
Birth cohorts, 33
Black box, 293

Bloomberg, Michael, 58
Bohn, Angela, 192
Boolean search logic, 347, 348 (exhibit)
Booth, Wayne C., 330
Boston McKinney Project, 300 (exhibit)
Broken windows theory, 21
Brooks, Clem, 276–277
Brown v. Topeka Board of Education, 21
Buchta, Christian, 192
Buffalo Creek, West Virginia, 28
Bureau of Labor Statistics (BLS), 189
Bush, George W., 31 (exhibit), 301

Campbell, Donald T., 310
Campbell Collaboration, 291
Careers and research
    consultant, 291 (box)
    graduate program faculty, 276 (box)
    policy work company, 188 (box)–189 (box)
    principal scientist, 80 (box)
    public health research, 328 (box)
    qualitative data analysis program, 245 (box)
    research assistant, 11 (box), 203 (box)
    research compliance specialist, 60 (box)
    research interest development, 36 (box)
    senior research assistant, 119 (box)
    sociologist, 103 (box)
    survey research company, 146 (box)
Carlisle Indian School, 246 (exhibit)
Case-oriented research, 272
Casey, Mary Anne, 221, 291 (box)
Categorizing and coding, qualitative data analysis, 236 (exhibit), 237 (exhibit)
Causal effect, 112
Causal invalidity
    contamination and, 127
    endogenous change and, 126, 326
    history effect and, 126–127
    noncomparable groups and, 124–125

regression effect and, 126
treatment misidentification and, 127–128
See also Generalizability
Causal validity (internal validity), 12–14 (exhibit)
threats to, 124 (exhibit)–128
Causation
association, 112–113
causal explanation in, 112
causal network model, 238 (exhibit)
context, 115
criteria, 112–115
mechanism, 115, 205
nonspuriousness, 113–115, 114 (exhibit)
quasi-experimental design and, 119–123, 121 (exhibit)
time order, 113, 117, 123
true experiment and, 116 (exhibit)–118
Cell phones, 3, 151, 152, 157–158, 246
Census, 94, 187, 189
Center for Survey Research, 151
Centers for Disease Control and Prevention (CDC), 191–192
Central tendency, measures of, 169, 177–178
Certificate of Confidentiality, 56–57
Ceteris paribus, 112
Chaffee, Benjamin W., 179 (box)
Chambliss, Dan, 26–27, 32, 75, 200–201, 206, 207, 208, 212, 213 (exhibit), 222, 251, 263–264
Champions: The Making of Olympic Swimmers (Chambliss), 26–27
Change, resistance to, 6
Charlotte-Mecklenburg Police Department (CMPD), 27 (box)
Children
child abuse prevention, 25
informed consent and, 55
Chirac, Jacques, 58
Chi-square, 185
Christakis, Nicholas, 55
Closed-ended (fixed-choice) question, 72–73, 143, 153, 316
Cluster, defining, 100
Cluster sampling, 99–101 (exhibit)
Cocaine use, 172–173 (exhibit), 174 (exhibit), 211 (exhibit)

Code of Ethics (ASA), 47–48, 131–132, 334
Code of the Street (Anderson), 202–203
Coding
in content analysis, 268–269
in content analysis, sample form, 270 (exhibit)–271 (exhibit)
in qualitative data analysis, 72, 236 (exhibit), 237 (exhibit)
Cognitive interview, 145
Cohen, Alison K., 179 (box)
Cohen, Gary, 13–14
Cohen, Susan, 123
Cohort, 32–33
Cohort design, 32–33
Coleman, James, 33
Collins, Randall, 262, 263 (exhibit)
Colomb, Gregory G., 330
Coming of Age in Samoa (Mead), 222
Community health workers (CHWs) case study, 322–328
Community Mental Health Act Amendment (1975), 291
Comparative methods, 275–283
cautions, 282–283
cross-sectional, 276–280, 277 (exhibit)–279 (exhibit)
longitudinal, 281 (exhibit)–282
Comparison
of levels of measurement, 77 (exhibit), 79–80 (exhibit)
of research designs, 315–317, 316 (exhibit)
Comparison group, 116
Compensatory rivalry (John Henry effect), 127
Competitive swimmer study, 26–27, 207
Complete observation, 206–207
Complete (covert) participation, 207–208
Computer-assisted personal interview (CAPI), 153
Computer-assisted qualitative data analysis, 251–254, 252 (exhibit)–253 (exhibit)
Computer-mediated counseling, 12–14 (exhibit)
Concept, 66–67
Conceptualization, 67–68
in qualitative data analysis, 233, 235–236
Confidentiality, 56–57, 310
in qualitative research, 222

in survey research, 150, 159
Connor, Jennie, 155
Consent, informed, 51–55, 51 (exhibit)–54 (exhibit), 132
Constants, 68–69
Construction of questions, 72–75
Construct validity, 81–82
Consumer Expenditure Survey, 187
Contact summary form, 234 (exhibit)
Contamination, 127
Content analysis, 72, 265–271
coding procedures, 268–269, 270 (exhibit)–271 (exhibit)
flowchart of, 266 (exhibit)–267 (exhibit)
identifying population of documents/textual sources, 265
sampling in, 267–268 (exhibit)
statistical analyses development, 269
unit of analysis in, 265
Content validity, 266 (exhibit)
Context, 10, 115, 123
Context effects, 146
Contingent question, 143
Contrived observation, 264–265
Control group, 116–117
Convenience (availability) sampling, 103–104, 156
Conversation analysis, 243–244 (exhibit)
Correlation, 112–114
Correlation analysis, 186
Cost-benefit analysis, 307–309, 308 (exhibit)
Cost-effectiveness analysis, 307
Council of European Social Science Data Archives (CESSDA), 193
Cover letter, for mailed survey, 149 (exhibit)–150
Crabtree, Benjamin F., 231
Crime and criminal justice system
causation and, 114 (box)
context effect on crime, 115
deterrence theory and, 22, 23 (exhibit), 24, 26
divergent research results on, 76
incarceration alternatives study, 299
juvenile court workers study, 249–250 (exhibit)
juvenile justice processing study, 247–248

neighborhood police officer (NPO) study, 233, 239
physical disorder effect on crime study, 214–215, 216 (exhibit)–217 (exhibit)
police interaction with public study, 75
police reform study, 21
underreporting of c, 140
women's prison and education study, 306–307
See also Domestic violence
Crime Control and Safe Streets Act, 310
Criterion validity, 81, 82
Critical case, 210
Cross-population generalizability (external validity), 12, 13 (exhibit), 94, 129–130
Cross-sectional comparative research, 276–280, 277 (exhibit)–279 (exhibit)
Cross-sectional research design, 29, 30 (exhibit)
repeated, 30 (exhibit), 31 (exhibit)
Cross-tabulation (crosstab), 182–186
reading table in, 182 (exhibit)–185, 183 (exhibit)–184 (exhibit)
trivariate, 185–186 (exhibit)
Current Population Survey (CPS), 187, 188
Cutshall, Charles R., 262
Cyberethnography, 203–204

Data, using available, 71–72
Databases, 268, 346–347
Data cleaning, 168
Debriefing, 49
Decay, instrument, 125
Deception, 55–56, 131–132
Decker, Scott, 252, 254
Deductive research, 24–26
conceptualization in, 233, 235–236
Demoralization, 127
Denzin, Norman K., 254–255
Dependent variable, 26
posttests and pretests, 117
sampling on, 92
See also Deductive research
Depression, 74 (exhibit)

Descriptive research, 8 (exhibit), 28
Descriptive statistics, 168
Descriptive theory, 294, 295
Deselection, 125
Design, research, 28–37
    cohort, 32–33
    comparing, 30 (exhibit)
    cross-sectional, 29, 30 (exhibit)
    evaluation (*see* Evaluation research)
    longitudinal, 29–33, 30 (exhibit)
    panel, 30 (exhibit), 32
    quasi-experimental, 119–123, 121 (exhibit)
    questionnaire, 144–147 (exhibit)
    survey (*see* Survey research)
    true experiment, 116 (exhibit)–118
    units and levels of analysis, 29, 33–37, 35 (exhibit)
Desmond, Matt, 203, 205
Deterrence theory, 22, 23 (exhibit), 24, 26
Detroit Area Study, 142
Deviant case, 210
Dewey, Thomas F., 97
Diamond, Timothy, 205, 207, 210, 296
Differential attrition (mortality), 125
Dill, Karen, 265, 268, 269
Dillman, Don A., 142
Direction of association, 26. *See also* Deductive research
Direct observation, 75
Disasters, 28, 248–249
Disproportionate stratified sampling, 102–103
Distribution of benefits, 132
Divergence between measures, 76
Diversity of population, 94
Documentation in qualitative data analysis, 233
Domestic violence
    appropriate action on, 27 (box), 60–61
    causal effects and, 112
    descriptive research on, 27 (exhibit), 28
    generalizability of research on, 129
    literature search for, 319–321

needs assessment for, 301, 302 (exhibit)–303 (exhibit)
research circle and, 26, 27 (exhibit)
selection bias in studies of, 125
theory and, 23 (exhibit), 24
See also Minneapolis Domestic Violence Experiment
Double-barreled question, 141
Double-blind procedure, 127
Double negative, 141
Drake, Robert E., 128, 306
Drug abuse
drug sales study, 252, 254
mentally ill homeless persons and, 250 (exhibit)–251 (exhibit)
Drug Abuse Resistance Education (D.A.R.E.), 289–290, 297, 303–304 (box)
Duckworth, Kenneth, 265, 268, 269
Duneier, Mitchell, 34
Durkheim, Émile, 33–34, 187, 272

Earthquake response study, 248
Ecological fallacy, 34
Ecoterrorism, 56
Education
effect on Internet usage, 3, 4 (exhibit), 155
effect on voting, 179 (box), 183–184 (exhibit), 185 (exhibit), 186 (exhibit)
highest year of school completed, 172 (exhibit)
masculinity/bullying/academics study, 243
private *vs.* public schools study, 3, 33, 433
socioeconomic status effect on, 114 (exhibit)–115
years of education completed, 177 (exhibit)
Effect of external events (history effect), 126–127
Efficiency analysis, 307–309, 308 (exhibit)
Einstein, Albert, 66
Ekman, Paul, 263, 264 (exhibit)
Elaboration analysis, 185
Electronic survey, 154–157 (exhibit), 158 (exhibit), 159
Elements of sample, 92
Elliott, Andrew, 130
Ellis, Carolyn, 58

E-mail and social interaction, 3
E-mail survey, 154, 155
Emic focus, 230
Employment studies
call back outcomes study, 129
guaranteeing jobs to poor people, 297–298 (exhibit)
services for persons with severe mental disorders, 306 (exhibit)
young homeless persons, 95 (box)
Endogenous change, 126, 326
Equal probability of selection method (EPSEM), 99
Erikson, Kai, 28, 208
Errors
ecological fallacy, 34
everyday, 4–6
reductionism, 37
Ethics, 43–61
case example, 43–45 (exhibit), 44 (exhibit)
confidentiality and, 56–57, 150, 159, 222, 310
deception and, 55–56, 131–132
evaluation research and, 309–310
historical background of, 46 (exhibit)–47
honesty/openness and, 57–58
informed consent and, 51–55, 51 (exhibit)–54 (exhibit), 132
institutional review board and, 47, 309–310
plagiarism and, 334
principles of, 47–61
privacy and, 56–57, 310
protecting research subjects and, 48–57
qualitative data analysis and, 254
qualitative research and, 208
quantitative data analysis and 192–193
survey research and, 159–160
unobtrusive measures and, 283
Ethnography, 202–203
Ethnomethodology, 205
Etic focus, 230
European Social Survey (ESS), 277–278
Evaluability assessment, 300–301, 302 (exhibit)–303 (exhibit)
Evaluation research, 289–310

black box or program theory, 293–295
characteristics of, 292–293
defining, 10
design alternatives, 293–298
efficiency analysis, 307–309, 308 (exhibit)
ethical issues in, 309–310
evaluability assessment, 300–301, 302 (exhibit)–303 (exhibit)
focus of, 299–309
history of, 290–291
hypothesis, 26
impact analysis, 305–307, 306 (exhibit)
model of, 292 (exhibit)
needs assessment, 299–300 (exhibit), 301, 302 (exhibit)–303 (exhibit)
process evaluation, 301, 303–305, 304 (exhibit)
quantitative or qualitative methods, 296–297
researcher or stakeholder orientation, 295–296
simple or complex outcomes in, 297–298 (exhibit)
Event cohorts, 33
Event-structure analysis, 273, 274 (exhibit)
Everything in Its Path (Erikson), 28
Exhaustive, 73
Expectancies of experiment staff (self-fulfilling prophecy), 127
Expense and attrition in panel design, 32
Experimental group, 116–117
Experimental research
comparison to other designs, 316 (exhibit)
field experiment, 128–129
protecting subjects of, 48–49
quasi-experimental designs, 119–123, 121 (exhibit)
threats to validity in, 123–128
true experiments, 116 (exhibit)–118
validity and, 305–306
Explanatory research, 10, 26
Exploratory research, 8–9, 200, 205
Ex post facto control group design, 120, 123
External validity (cross-population generalizability), 12, 13 (exhibit), 129–130
Extraneous variable, 185 (exhibit)
Exum, M. Lyn, 27 (box)

Facebook, 1, 2, 55, 156, 190, 192, 247, 264
Face validity, 81, 266 (exhibit)
Fallon, Kathleen, 282 (box)
Family Justice Center (FJC), 301
Fatigue, subject, 32
Federal Policy for the Protection of Human Subjects, 47
Feedback, 292
Fence-sitters, 144
Fennell, Mary, 36 (box)
Fenno, Richard, 208
Ferguson, Kristin, 95 (box)
Ferguson, Ross, 114 (box)
Field, Jennifer, 260
Field experiment, 128–129
Field notes, 212, 213 (exhibit)
Field research
    choosing role in, 206–208
    complete observation role, 206–207
    complete participation in, 207–208
    developing and maintaining relationships in, 209
    entering field, 208
    managing personal dimensions in, 212–214 (exhibit)
    mixed, 207
    note taking in, 212, 213 (exhibit)
    sampling people and events in, 210 (exhibit)–212, 211 (exhibit)
Filter question, 143 (exhibit), 144
Fine, Gary Allen, 201
Fink, Arlene, 350
Firefighter study, 203, 205
Floaters, 143, 144
Focus groups, 145, 220–221 (exhibit)
Formative evaluation, 305
Fourth Generation Evaluation (Guba & Lincoln), 295
Fowler, Floyd J., Jr., 146 (box)
Fox, Nick, 222
Franklin, Mark, 278–280, 279 (exhibit)
Freedom of Information Act (FOIA), 283
Frenkel-Brunswick, Else, 82
Frequency distribution, 173–175 (exhibit), 176 (exhibit)

Frequency polygon, 172–173 (exhibit), 174 (exhibit)
Freshman, Audrey, 156
Freud, Sigmund, 94, 222
Friday, Paul C., 27 (box)
Frieden, Thomas, 58
Front matter (reports), 331
Furr-Holden, C. Debra, 57 (box)

Gallo, Robert, 58
Gallup poll, 12, 31 (exhibit), 93, 97, 98 (exhibit), 141
Gamma, 184–185
Gang studies, 67 (box), 207, 211–212, 239, 297
Gatekeeper, 209
Gaubatz, Kurt Taylor, 276 (exhibit)
Gender
    affect on voting, 184 (exhibit)
    gender categories in social life study, 240
    representation in government study, 282 (box)
    See also Domestic violence
Gender equality, 280
Generalizability
    cross-population, 12, 13 (exhibit), 94, 129–130
    evaluating, 93–94
    overgeneralization and, 4, 5, 6 (exhibit), 7
    sample, 12, 13 (exhibit), 128–129
    threats to, 128–130
General Social Survey (GSS), 71, 140, 187
    infidelity study, 153, 175
    neighboring study, 8 (exhibit)
    voting/political participation, 168, 169 (exhibit), 175 (exhibit)
    years of education study, 171, 172 (exhibit)
Gerth, Matthias, 265, 268
Gibson, David R., 243–244 (exhibit)
Gilbert, Dennis, 139
Gillespie, Chris, 268
Gilligan, Carol, 242–243
Ginsberg, Jeremy, 191–192
Glaser, Barney G., 211
Glossary, 338–344

Goal-free evaluation, 295–296
Goffman, Alice, 203
Goffman, Erving, 33, 207, 222
Going native, 208
Goode, Erich, 58
Google, 191–192, 351 (exhibit)
Google Scholar, 346–347
Gosling, Samuel D., 55
Goulet, Lauren Sessions, 2
Government Performance and Results Act (1993), 291
Graffiti study, 262
Graphs, 171 (exhibit)–173 (exhibit)
Gray, Andrew, 155
Griffin, Larry, 272–273
Grounded theory, 244–245
Group-administered survey, 150, 160
Group unit of analysis, 29, 36
Guba, Egon, 295
Gubrium, Jaber F., 205
Guns
    gun control study, 321
    gun violence survey, 155
    youth and guns survey, 152 (exhibit)
Gupta, Neeti, 7 (box)

Habits of the Heart (Bellah), 217–218, 219
Hampton, Keith, 2, 7 (box)
Handbook of Research Design and Social Measurement, 145
Haphazard sampling, 96
Hartman, Jennifer L., 27 (box)
Harvey, Daina, 209 (box)
Hawthorne effect, 128
Health Care Finance Administration (HCFA), 305
Health Insurance Portability and Accountability Act (HIPAA), 56
Health Research Extension Act (1985), 309–310
Heat Wave (Klinenberg), 34
Hermeneutic perspective, 230
Histogram, 171–172 (exhibit)
Historical research, 272–275

History effect (effect of external events), 126–127
Hitler, Adolf, 82
Hoffer, Thomas, 33
Holistic research, 272
Holstein, James A., 205
Homelessness
developing as research interest, 36 (box)
employment and young homeless persons, 95 (box)
mentally ill homeless persons and drug abuse, 250 (exhibit)–251 (exhibit)
mentally ill homeless persons housing choice, 300 (exhibit)
mentally ill homeless persons program report, 331, 332 (exhibit)
Homosexuality, 55
Honesty, 57–58
Hornik, Kurt, 192
Houtzager, Peter, 276
How College Works (Chambliss & Takacs), 32
How to Lie With Statistics (Huff), 193
Hoyle, Carolyn, 319–320
Huberman, A. Michael, 233, 254
Huff, Darrell, 193
Human Relations Area Files (HRAF), 190
Humphreys, Laud, 55, 207
Hunt, Dana, 80 (box), 301
Hurricane response study, 248–249
HyperRESEARCH, 251–252 (exhibit)
Hypothesis, 25–26. *See also* Deductive research

Illogical reasoning, 4, 5–6, 7
Immigrant studies, 277–278 (exhibit), 296–297
Impact analysis (impact evaluation/summative evaluation), 305–307, 306 (exhibit)
Inaccurate (selective) observation, 4, 5, 6 (exhibit), 7
Incarceration alternatives study, 299
In Defense of Food (Pollan), 262
Independent variable, 26. *See also* Deductive research
Index, 73–75, 74 (exhibit)
Individual matching, 120
Individual unit of analysis, 29

Inductive reasoning, 27
Inductive research, 26–28, 230
    conceptualization in, 233, 235–236
Inferential statistics, 168
Infidelity study, 153, 175
Informed consent, 51–55, 51 (exhibit)–54 (exhibit), 132, 222
In-person interviews, 153, 158 (exhibit), 159–160
Inputs, defining, 292
Institutional racism, 21
Institutional review board (IRB), 47, 309–310
Instrument decay, 125
Integrative approaches (to evaluation), 296
Intensive (depth) interviewing, 204, 215, 217–218
    asking questions/recording answers in, 218–219
    establishing/maintaining partnership in, 218
    saturation point and, 218 (exhibit)
Interactive voice response (IVR), 153
Interitem reliability (internal consistency), 83
Internal consistency (interitem reliability), 83
Internal validity. *See* Causal validity (internal validity)
Internet
    cyberethnography, 203–204
    data collection via, 168, 170 (exhibit)
    effect of education on usage of, 3, 4 (exhibit), 155
    ethics and research on, 222
    inequality in access to, 9
    interviewing via, 219–220
    netnography, 203–204
    social networking, 1–2
    social ties survey and, 2–4
    surveys, 153–157 (exhibit)
    web literature searches, 351–353
    web survey, 154–156, 158 (exhibit), 159
    wireless access to, 3, 7 (box)
Interobserver reliability, 83
Interpretation of Dreams (Freud), 222
Interpretive questions, 145
Interquartile range, 180
Inter-University Consortium for Political and Social Research (ICPSR), 71,

187, 188 (exhibit), 189–190, 193
Interval level of measurement, 77 (exhibit), 78
Interview guide, 147 (exhibit)
Interviewing. *See* Intensive (depth) interviewing
Interview schedule, 144
Invalidity. *See* Causal invalidity; Validity
Irvine, Leslie, 222

Jankowski, Martin Sanchez, 207, 211–212
Jelly's Bar study, 202–203, 231, 235, 240–241
Jenson, Jakob, 113
Jerolmack, Colin, 203
Jihong "Solomon" Zhao, 67 (box)
John Henry effect (compensatory rivalry), 127
Jottings, 212
Justice, 47 (exhibit)
Juveniles
    court workers study, 249–250 (exhibit)
    juvenile justice processing study, 247–248
    See also Crime and criminal justice system

Kale-Lostuvali, Elif, 248
Kato, Yuki, 209 (box)
Kedrosky, Paul, 260
Kelling, George, 21
Kenny, Kristen, 60 (box)
Kerr, Barbara, 13–14
Key informant, 105, 209
Kilgore, Sally, 33
Kinsey, Alfred, 21
Klinenberg, Eric, 34, 204
Klofas, John M., 262
Koppel, Ross, 103 (box)
Kozinets, Robert V., 204
Krueger, Richard, 220, 221
Kypri, Kypros, 155

Labeling theory, 23 (exhibit), 24
Labov, William, 264–265

Lacey, Marc, 57 (box)
Landers, Ann, 4
Latino identity, 68
Laub, John H., 115
LeBlanc, Jessica, 11 (box)
Ledford, Gerald, Jr., 123
Lee, Mei Hsien, 75
Lelieveldt, Herman, 10
Level(s) of measurement
    comparison among, 77 (exhibit), 79–80 (exhibit)
    interval, 77 (exhibit), 78
    nominal, 76–78, 77 (exhibit)
    ordinal, 77 (exhibit), 78, 79 (exhibit)
    ratio, 77 (exhibit), 78–79
Levels of analysis, 29, 33–37, 35 (exhibit)
Levinson, Daniel, 82
LexisNexis database, 268
Lieberson, Stan, 117
Lincoln, Yvonna S., 295
Ling, Rich, 2
Ling Ren, 67 (box)
Linguistic field experiment, 264–265
LinkedIn, 2
Listening, active, 217
Literature search, 24, 317–321, 345–353
    conducting, 346–350
    “Night as Frontier” case study, 317–319
    preparing for, 346
    questions to ask about research articles, 317, 318 (exhibit)
    web searches, 351–353
    “When Does Arrest Matter” case study, 319–321
    writing review, 349–350 (exhibit)
Livingston, Jennifer, 249
Longitudinal comparative research, 281 (exhibit)–282
Longitudinal research design, 29–33, 30 (exhibit)
Lord, Vivian B., 27 (box)
Luker, Kristin, 222
Lynching, 273, 274 (exhibit)

Maccio, Elaine, 95 (box)
Mailed (self-administered) survey, 148–150
advantages/disadvantages of, 158 (exhibit)
anonymity and, 160
cover letter for, 149 (exhibit)–150
response rates of, 157
Mair, Patrick, 192
Makarios, Matthew, 321 (box)
Manza, Jeff, 276–277
Margolis, Eric, 246 (exhibit)
Marshall, Gary, 55–56
Marshall, Ineke Haen, 67 (box)
Masculinity/bullying/academics study, 243
Matching, 117
aggregate, 121
individual, 120
Matrix, 74 (exhibit), 236 (exhibit)
Maturation, 126
McCarter, Susan, 247–248
McLellan, A. Thomas, 82
Mead, Margaret, 222
Mean, 178, 179 (exhibit)
Measurement
central tendency, 169, 177–178
combined operations, 75–76
comparison of levels of, 77 (exhibit), 79–80 (exhibit)
interval level of, 77 (exhibit), 78
nominal level of, 76–78, 77 (exhibit)
ordinal level of, 77 (exhibit), 78, 79 (exhibit)
ratio level of, 77 (exhibit), 78–79
unobtrusive measures (*see* Unobtrusive measures)
validity, 12, 81–82
variation, 178, 180 (exhibit)–181
Measure of association, 184–185
Mecca, Laurel Person, 245 (box)
Mechanism, 115, 205
Media effects, 113
Median, 177, 178, 179 (exhibit)
Medicaid, 118, 210, 305

Medical students study, 235
MEDLINE, 346
Melbin, Murray, 317–319
Mellon Project, 220
Mental health system effectiveness study, 249
Miles, Matthew B., 233, 254
Milgram, Stanley, 43–45 (exhibit), 44 (exhibit), 48–49, 50, 55, 57–58, 59, 60, 131
Military suicides, 247
Miller, Susan, 233, 239
Miller, William L., 231
Mills, Judson, 56, 131
Minneapolis Domestic Violence Experiment, 27 (exhibit), 28, 60–61, 125, 297. *See also* Domestic violence
Mixed methods, 246–251, 250 (exhibit)
Mixed participation/observation, 207
Mobile phones. *See* cell phones
Mode (probability average), 177
Montagnier, Luc, 58
Mooney, Christopher, 75
Moore, Barrington, Jr., 272
Moore, Spencer, 248–249
Moral development theory, 242–243
Morgan, Philip, 275
Morrill, Calvin, 229, 231, 235, 240–241
Mortality (differential attrition), 125
Mullins, Carolyn J., 330
Multiple group before-and-after design, 122 (exhibit)–123
Munafo, Marcus, 329
Muslim America project, 139
    pilot study, 145
    response rate, 152
Mutually exclusive, 73
Myrdal, Gunnar, 21
MySpace, 2, 55

Narrative analysis, 241–243, 242 (exhibit)
Narrative explanations, 272–273
National Archive of Criminal Justice Data, 187

National Cancer Institute (NCI), 322–328
National Geographic Society, 12, 155
National Institutes of Health (NIH), 322, 325
National Opinion Research Center (NORC), 71, 187
National Roadside Survey, 57 (box)
Needleman, Carolyn, 249–250 (exhibit)
Needs assessment
    domestic violence, 301, 302 (exhibit)–303 (exhibit)
    housing for homeless mentally ill persons, 299–300 (exhibit)
Neighborhood police officer (NPO) study, 233, 239
Neighboring, 8 (exhibit)
Nesta, Daniela, 130
Netnography (cyberethnography/virtual ethnography), 203–204
Neuendorf, Kimberly, 265, 268 (exhibit), 269
Newbury, Darren, 246
Newman, Katherine S., 92
Ngrams, 191 (exhibit), 192
“Night as Frontier” case study, 317–319
Ni “Phil” He, 67 (box)
Nixon, Richard, 141
Nominal level of measurement, 76–78, 77 (exhibit)
Noncomparable groups, 124–125
Nonequivalent control group design, 120–121 (exhibit)
Nonprobability sampling methods, 103–105
Nonrandom sampling, 267
Nonresponse, 96, 151 (exhibit)–152
Nonspuriousness, 113–115, 114 (exhibit)
Normal distribution, 181 (exhibit)
Note taking, 212, 213 (exhibit)
Numbers, table of random, 354–358
Nuremberg war crime trials, 46
Nursing home study, 296
NVivo, 251–252, 253 (exhibit)

Obedience experiments (Milgram’s), 43–45 (exhibit), 44 (exhibit), 48–49, 50, 55, 57–58, 59, 60, 131
Observation
    as unobtrusive measure, 263–264
    complete, 206–207

contrived, 264–265
direct, 75
selective (inaccurate), 4, 5, 6 (exhibit), 7
systematic, 214–215, 216 (exhibit)
See also Participant observation
Office for Protection From Research Risks, National Institutes of Health, 47
Olympic-level competitive swimmer study, 26–27, 207
Omnibus survey, 140
Online data collection, 168, 170 (exhibit)
Online interviewing, 219–220
Online research, ethics and, 222
Online social networking, 1–2
Open-ended questions, 73, 144, 153, 249. *See also* Intensive (depth) interview
Openness, 57–58
Operation, defining, 69
Operationalization, 69–71, 70 (exhibit), 79
Optical illusion, 5, 6 (exhibit)
Oral history, 273, 275
Ordinal level of measurement, 77 (exhibit), 78, 79 (exhibit)
Organizational loyalty, 8–9
Organizing and writing reports. *See* Writing and organizing reports
Outcomes
defining, 292
employment call back study and, 129
evaluation research and, 297–298 (exhibit)
influence of context on adolescent, 10
Outlier, 180
Outputs, defining, 292
Overgeneralization, 4, 5, 6 (exhibit), 7. *See also* Generalizability

Pager, Devah, 129
Pagnini, Deanna, 275
Panel design, 30 (exhibit), 32
anonymity and, 159
repeated measures, 123
Parameter, 168
Participant observation, 205–209
choosing role in, 206–208

complete observation, 206–207
complete participation, 207–208
developing/maintaining relationships in, 209
entering field, 208
managing personal dimensions in, 212–214 (exhibit)
mixed, 207
note taking in, 212, 213 (exhibit)
observational continuum, 206 (exhibit)
sampling people/events in, 210 (exhibit)–212, 211 (exhibit)
Participation shifts (P-shift), 244 (exhibit)
Participatory research, 295
Passidomo, Catarina, 209 (box)
Paternoster, Ray, 321
Patton, Michael Quinn, 230
Percentage, 173–174
Periodicity, 99, 100 (exhibit)
Perry, Gina, 57–58
Phillips, David P., 122
Phoenix, Ann, 243
Phone survey
advantages/disadvantages of, 158 (exhibit)
anonymity and, 160
interactive voice response technology, 153
reaching sample units in, 150–151
response rates, 157–158
Photographic data, 262, 263 (exhibit)
Physical disorder effect on crime study, 214–215, 216 (exhibit)–217 (exhibit)
Physical traces, 260–261 (exhibit)
Piliavin, Irving, 56
Piliavin, Jane Allyn, 56
Pilot study, 145, 322
Pipher, Mary, 125
Placebo effect, 127–128
Plagiarism, 334
Police reform, 21
Political participation, 10, 168, 169 (exhibit), 175 (exhibit), 278–279 (exhibit), 286 (exhibit)
Pollan, Michael, 262

Pollio, David, 95 (box)
Polls, 72
    Gallup poll, 12, 31 (exhibit), 93, 97, 98 (exhibit), 141
    push, 159
Population, 91
    diversity of, 94
    target, 94
    vulnerable, 55, 98, 310
Posttest, 117, 120, 123
    endogenous change and, 126
Poverty, 26, 69, 70 (exhibit), 113, 115, 298
Pratt, Travis, 321 (box)
Prescriptive theory, 294–295
Presentation of Self in Everyday Life (Goffman), 222
President's Family Justice Center (FJC), 301
Pretest, 117, 120, 123
endogenous change and, 126
Price, Richard H., 129
Primary deviance, 24
Primary sampling units, 93 (exhibit)
Prime time network television study, 265, 268 (exhibit), 269
Prior research, 319–321
Prison simulation study (Zimbardo's), 49, 50 (exhibit)
Privacy, 56–57, 310
Private schooling, effect on achievement scores, 33. *See also* Education
Probability of selection, 96
Probability sampling method, 96–103
Process analysis, 128
Process evaluation, 301, 303–305, 304 (exhibit)
Program process, 292
Program theory, 294 (exhibit)–295
Progressive focusing, 230–231
Project New Hope, 297–298 (exhibit)
Proportionate stratified sampling, 101
Proposals, research. *See* Research proposals
Protection of research subjects, 48–57
PsycARTICLES, 346
PsycINFO, 346
Public schooling, effect on achievement scores, 33. *See also* Education

PubMed, 346
Purcell, Kristen, 2
Purposive sampling, 105
Push polls, 159
Putnam, Robert, 8, 10, 12

Qualitative data analysis, 229
alternatives in, 241–246
as art, 231–232
authenticating conclusions, 237, 239
checklist matrix, 236 (exhibit)
coding and categorizing in, 236 (exhibit), 237 (exhibit)
compared to quantitative data analysis, 232
computer-assisted, 251–254, 252 (exhibit)–253 (exhibit)
conceptualization in, 233, 235–236
contact summary form, 234 (exhibit)
conversation analysis, 243–244 (exhibit)
distinctive features of, 230–232
documentation, 233
ethical issues in, 254
examining relationships/displaying data in, 237 (exhibit), 238 (exhibit)
grounded theory, 244–245
historical and comparative research, 272–275
mixed methods, 246–251, 250 (exhibit)
narrative analysis, 241–243, 242 (exhibit)
reflexivity in, 240–241
tacit knowledge in, 239
techniques, 232–233
visual sociology, 245–246 (exhibit)
Qualitative methods, 200–223
case study, 200–201
comparison to other designs, 316 (exhibit), 317
ethical issues in, 208
ethnography, 202–203
ethnomethodology, 205
evaluation research and, 296–297
focus groups, 145, 220–221 (exhibit)
intensive interviews, 204, 215, 217–218 (exhibit)
netnography, 203–204

participant observation (*see* Participant observation)
Quane, James, 10
Quantitative data analysis, 167–194
compared to qualitative data analysis, 232
cross-tabulation, 182–186, 182 (exhibit)–185 (exhibit)
ethical, 193–194
evaluation research and, 296
frequency distributions, 173–175 (exhibit), 176 (exhibit)
graphs and, 171 (exhibit)–173 (exhibit)
options for displaying distributions in, 169, 171–175
options for summarizing distributions, 175, 177–181
preparing data for, 168, 170 (exhibit)
secondary data analysis, 186–190
Quartiles, 180
Quasi-experimental design, 119–123, 121 (exhibit)
Questionnaire, 144–147
attractiveness/ease of use of, 147 (exhibit)
Questions
allowing for disagreement, 142
allowing for uncertainty, 143–144
asking and recording answers to interview, 218–219
building on existing instruments, 144–145
clear, 141
closed-ended (fixed-choice), 72–73, 143, 153, 316
constructing, 72–75
contingent, 143
double-barreled, 141
exhaustive/mutually exclusive response categories, 73, 144
indexes and scales, 73–75, 74 (exhibit)
interpretive, 145
maintaining consistent focus, 145
minimizing bias in, 141–142
open-ended, 73, 144, 153, 249
ordering of, 146
refining and testing, 145
respondent competency and, 142–143
single questions, 72–73
social research, 22–23
writing survey, 141–144

Quillian, Lincoln, 129
Quota sampling, 104 (exhibit)

Radin, Charles A.
Rainie, Lee, 2
Rampage: The Social Roots of School Shootings (Newman), 92
Random assignment (randomization), 117, 118 (exhibit)
Random digit dialing (RDD), 99, 150–151
Random number table, 99, 354–358
Random sampling, 96
    random assignment *vs.*, 118 (exhibit)
    simple, 99, 267
    stratified, 101–103, 102 (exhibit), 267
    systematic, 99, 100 (exhibit)
Range, 180
Rankin, Bruce, 10
Ratio level of measurement, 77 (exhibit), 78–79
Rational choice theory, 23 (exhibit)
Raudenbush, Stephen, 214–215
Reactive effects, 206–207
Reactive methods, 259
Reagan, Ronald, 58
Reasoning
    illogical, 4, 5–6, 7
    inductive, 27
Reductionist fallacy (reductionism), 37
Reflexivity, 240–241
Reform, police, 21
Regime classification, Latin America, 281 (exhibit)–282
Regression effect, 126
Reiss, Albert J., Jr., 75
Reliability, 82–84
    achieving validity and, 83–84 (exhibit)
    alternate-forms, 83
    interitem, 83
    interobserver, 83
    split-halves, 83
    test-retest, 83
Repeated measures panel design, 123

Reporting research, 328–334
plagiarism and, 334
writing and organizing report, 329–334
Representative sample, 94, 95 (exhibit), 97
Research. *See* Social science research
Research circle, 24, 25 (exhibit)
domestic violence and, 26, 27 (exhibit)
Research proposals, 321–328
checklist of decisions for, 327 (exhibit)
community health workers case study, 322–328
sample proposal, 323 (exhibit)–324 (exhibit)
sections included in, 321–322
Resistance to change, 6
Respect for persons, 47 (exhibit)
Response rates, 96, 151 (exhibit)–152, 153, 154, 155, 157–158
Reverse outlining, 330
Reviving Ophelia (Pipher), 125
Rhodes, William, 301
Rinehart, Jenny, 238 (box)–239 (box)
Ringwalt, Christopher L., 303–304
Rivalry, compensatory, 127
Roberts, Chris, 222
Rossi, Peter H., 304 (exhibit)–305
Rubin, Herbert, 105
Rubin, Irene, 105
Rueschemeyer, Dietrich, 281 (exhibit)–282

Sacks, Stanley, 309
Sample generalizability, 12, 13 (exhibit), 128–129
Sample/sampling
availability, 103–104
bias and, 96
census, 94
choosing method for, 96–105
cluster, 99–101 (exhibit)
components and population in, 92–93 (exhibit)
definition of sample, 92
disproportionate stratified, 102–103
generalizability of, 12, 13 (exhibit), 93–94

interval, 99
nonprobability methods, 103–105
nonrandom, 267
people and events, 210 (exhibit)–212, 211 (exhibit)
population diversity assessment, 94
primary sampling units, 93 (exhibit)
probability, 96–103
probability of selection in, 96
proportionate stratified, 101
purposive, 105
quota, 104 (exhibit)
representative sample, 94, 95 (exhibit), 97
secondary sampling units, 93 (exhibit)
simple random, 99, 267
snowball, 105, 156, 249
stratified random, 101–103, 102 (exhibit), 267
systematic random, 99, 100 (exhibit)
theoretical, 211 (exhibit)–212
vulnerable populations, 98
Sampling frame, 92
Sampling interval, 99
Sampling units, 92–93 (exhibit)
Sampson, Robert J., 115, 214–215
Sanders, Andrew, 319–320
Sanford, Nevitt, 82
Santiccioli, Jessica
Saturation point, 218 (exhibit)
Scale, 75
Scarce, Rik, 56
Schapira, Lidia, 322–328
School shootings, 92
Schorr, Lisbeth B., 297
Schuck, Amie, 304 (box)
Schutt, Russell K., 36 (box), 250 (exhibit)–251, 265, 268, 269, 300 (exhibit), 322–328, 331, 333 (exhibit)
Science, 7
Scientific failures, reporting of, 329
Secondary data analysis, 186–190
Secondary deviance, 24

Secondary sampling units, 93 (exhibit)
Segregation, 21, 240
Selection bias, 125
Selective distribution of benefits, 132
Selective (inaccurate) observations, 4, 5, 6 (exhibit), 7
Self-fulfilling prophecy (expectancies of experiment staff), 127
Self-managed work team study, 123
Self-reports, 262
Seniority cohorts, 33
Serendipitous, 27
Severe initiation experiment, 56, 131
Sexual attraction study, 130
Sexual Experiences Survey (SES), 238 (box)–239 (box)
Sexuality, 21
Sherman, Lawrence W., 26, 60–61, 112, 129, 132, 297
Shock experiments. *See* Obedience experiments (Milgram's)
Shootings, school, 92
Siegert, Gabriele, 265, 268
Silverman, Stephen J., 140
Simple random sampling, 99, 267
Single questions, 72–73
Skewness, 169, 171, 172, 175 (exhibit), 178
Skip pattern, 143 (exhibit)
Smith, Adam, 36–37
Smutylo, Terry, 307
Snowball sampling, 105, 156, 249
Snowden, Edward, 264
Social capital, 10, 12, 15, 192
Social health, 66
Social isolation, 3
Social networking
    Facebook and, 1, 2, 55, 156, 190, 192, 247, 264
    Twitter and, 191, 247
    YouTube and, 190–191, 262
    See also Internet
Social Networking Sites and Facebook Survey, 2
Social research question, 22–23
Social science, 7
Social science approaches (to evaluation), 295–296

Social Science Citation Index (SSCI), 347
Social science research, 12–14 (exhibit)
achieving valid results, 59
appropriate application of, 60–61
deductive, 24–26
descriptive, 8 (exhibit), 28
design, 28–37
errors in, 4–6
ethics and (*see* Ethics)
evaluation, 10
explanatory, 10
exploratory, 8–9
generalizability and, 12, 13 (exhibit), 93–94, 128–129
honesty/openness in, 57–58
hypothesis, 25–26
illogical reasoning in, 4, 5–6, 7
inductive, 26–28
in practice, 7–10
measurement validity, 12
overgeneralization in, 4, 5, 6 (exhibit), 7
proposing new, 321–328
protecting subjects of, 48–57
reporting, 328–334
research circle, 24, 25 (exhibit), 26
resistance to change in, 6
reviewing, 349–350 (exhibit)
selective/inaccurate observation in, 4, 5, 6 (exhibit), 7
strategy, 24–28
summarizing prior, 319–320
theory, 23 (exhibit)–24
units/levels of analysis, 29, 33–37, 35 (exhibit)
writing and organizing reports, 329–334
See also Subjects, research
SocINDEX, 346, 347
Socioeconomic status
effect on education performance, 114 (exhibit)–115
effect on voting, 182 (exhibit)–183, 184 (exhibit), 185 (exhibit), 186 (exhibit)
Internet usage and, 3, 155

operationalizing concept of, 70 (exhibit)
Sociological Abstracts, 346, 347, 350 (exhibit)
Sociology, visual, 245–246 (exhibit)
Solomon four-group design, 130 (exhibit)
Split-halves reliability, 83
Spurious, 113, 114 (exhibit), 123, 185 (exhibit)
Stake, Robert E., 233, 235
Stakeholder approaches (to evaluation), 293, 295, 296
Stald, Gitte, 2
Standard deviation, 180–181
Statistic, defining, 169
Statistical significance, 185
St. Jean, Pierre, 215, 217 (exhibit)
Strategy, social science research, 24–28
Stratified random sampling, 101–103, 102 (exhibit), 267
Straus, Murray, 61
Strauss, Anselm L., 211, 268, 269
Strauss, Jaine, 265
Street Corner Society (Whyte), 209
Strunk, William, Jr., 330
Student substance abuse, 331, 333 (exhibit)
Subject descriptors, 347, 349 (exhibit)
Subject fatigue, 32
Subjects, research
fatigue and, 32
informed consent and, 51–55, 51 (exhibit)–54 (exhibit), 132
protecting, 48–57, 130–132, 221–222
Suicide, 33–34, 187
before-and-after design, 121 (exhibit)
military suicides, 247
multiple group before-and-after design, 122 (exhibit)–123
Survey Documentation and Analysis (SDA), 187
Survey of Manufactures, 187
Survey research
comparison among survey designs, 147–148 (exhibit), 157–159, 158 (exhibit)
comparison to other designs, 316 (exhibit)
electronic, 154–157 (exhibit), 158 (exhibit), 159
ethical issues in, 159–160

group-administered, 150
in-person interviews, 153
interview guide, 147 (exhibit)
mailed, self-administered, 148–150, 157
online interviewing, 219–220
phone, 150–153
popularity of, 140
questionnaire design, 144–147 (exhibit)
writing questions for, 141–144
Swiss, Liam, 282 (box)
Symbolic interactionism, 23 (exhibit)
Synchronous online interviewing, 219–220
Systematic observation, 214–215, 216 (exhibit)
Systematic random sampling, 99, 100 (exhibit)

Table of random numbers, 354–358
Tacit knowledge, 239
Takacs, Chris, 32, 251
Target population, 94
Tearoom Trade (Humphreys), 55
Technology
ethical problems with, 55
See also Internet
Telephone survey. *See* Phone survey
Television study, prime time networks, 265, 268 (exhibit), 269
Testa, Maria, 249
Test-retest reliability, 83
Theoretical sampling, 211 (exhibit)–212
Theory
black box/program, 293–295
broken windows, 21
descriptive, 294, 295
deterrence, 22, 23 (exhibit), 24, 26
grounded theory, 244–245
labeling, 23 (exhibit), 24
moral development, 242–243
prescriptive, 294–295
program, 294 (exhibit)–295
rational choice, 23 (exhibit)

Theory-driven evaluation, 294
Therapeutic community (TC), 308–309
Thill, Kathryn, 265, 268, 269
Thompson, Sanna, 95 (box)
Thorne, Barrie, 213–214, 240
Tichenor, Veronica, 22
Time order, 113, 117, 123
Time series design, 123
Tipton, Steven, 219
Townsend, Meg, 301
Transitional Aid Research Project (TARP), 304 (exhibit)–305
Treatment misidentification, 127–128
Trend (repeated cross-sectional) design, 30 (exhibit), 31 (exhibit)
Triangulation, 76
Trivariate cross-tabulation, 185–186 (exhibit)
True experiment, 116 (exhibit)–118
Truman, Harry S., 97
Turabian, Kate L., 330
Tuskegee syphilis study, 46 (exhibit)
Twitter, 191, 247
Typical case, 210

Unimodal distribution, 177
Union literature study, 269, 270 (exhibit)–271(exhibit)
Units of analysis, 33–37, 35 (exhibit)
    group, 29, 36
    individual, 29
Unobtrusive measures, 259–283
    archives, 261–263 (exhibit)
    comparative methods, 275–283
    content analysis, 265–271
    contrived observation, 264–265
    creative sources of, 260–265
    ethical issues in, 283
    historical research, 272–275
    observation, 263–264
    physical traces, 260–261 (exhibit)
Urban gardening study, 209 (box)
U.S. Bureau of the Census, 94, 187, 189

U.S. Uniform Crime Reporting System, 275
Utilization-focused evaluation, 295

Validity, 11–14
achieving, 59
causal (internal validity), 12–14 (exhibit), 124 (exhibit)–128
construct, 81–82
contamination and, 127
content, 266 (exhibit)
criterion, 81, 82
endogenous change and, 126
experimental research and, 123–128, 305–306
external, 12, 13 (exhibit), 94, 129–130
face, 81, 266 (exhibit)
generalizability and, 12, 13 (exhibit)
history effect and, 126–127
interaction of testing/treatment and, 130
noncomparable groups and, 124–125
predictive, 342
reliability and, 83–84 (exhibit)
treatment misidentification and, 127–128
web surveys and, 156
See also Causal invalidity
Van Ryn, Michelle, 129
Van Winkle, Barrik, 252, 254
VanZile-Tamsen, Carol, 249
Variability, 169
Variable, 25–26
bimodal distribution of, 177
constants and, 68–69
cross-tabulation, 182 (exhibit)–186, 183 (exhibit)–184 (exhibit)
dependent, 26, 92, 117
exhaustive, 73
extraneous, 185 (exhibit)
independent, 26
mutually exclusive, 73
unimodal distribution of, 177
See also Deductive research
Variable-oriented research, 276

Variation, measures of, 178, 180 (exhibit)–181
Vinokur, Amiram D., 129
Violence against women study, 249
Violence: A Micro Sociological Theory (Collins), 262, 263 (exhibit)
Virtual ethnography, 203–204
Visual sociology, 245–246 (exhibit)
Viterna, Jocelyn, 282 (box)
Voluntary participation, 221, 222
Voting
    by family income, 179 (box), 182 (exhibit)–184 (exhibit)
    effect of education on, 179 (box)
    frequency distribution of voting, 175 (exhibit)
    voter participation study, 278–279 (exhibit), 286 (exhibit)
    voting/political participation study, 168, 169 (exhibit), 175 (exhibit)
Vulnerable populations, 55, 98, 310

Wageman, Ruth, 121
Webb, Eugene, 264
Webb, Vincent, 67 (box)
Weber, Max, 272, 282
Web searches, 351–353. *See also* Internet
Web survey, 154–156, 158 (exhibit), 159
Welfare state study, 276–277 (exhibit)
Welfare-to-work study, 293–294
Wells, William, 155
Westby, Ruth, 328 (box)
“When Does Arrest Matter” case study, 319–321
White, E. B., 330
Whyte, William F., 209, 222, 239
Williams, Joseph M., 330
Wilson, James Q., 21
Winiarek, Claire Wulf, 188 (box)–189 (box)
Wireless access, 3, 7 (box)
Withholding treatment, 49–50
Women’s prison and education study, 306–307
Women’s representation in government study, 282 (box)
Writing and organizing reports, 329–334
    applied reports, 331, 332 (exhibit)–333 (exhibit)
    back matter, 331

front matter, 331
journal article sections, 330–331 (exhibit)
reverse outlining and, 330
Writing skills study, 251

Xerox Corporation, 121 (exhibit)

Yankelovich, Daniel, 297
Yeater, Elizabeth, 238 (box)–239 (box)
Youth and guns survey, 152 (exhibit)
YouTube, 190–191, 262

Zaslow, Martha, 294
Zimbardo, Philip G., 49, 50 (exhibit), 55–56
Zogby International, 12
Zuckerberg, Mark, 1

SAGE researchmethods

The essential online tool for researchers from the world's leading methods publisher

Find exactly what you are looking for, from basic explanations to advanced discussion

More content and new features added this year!

Discover **Methods Lists**—methods readings suggested by other users

"I have never really seen anything like this product before, and I think it is really valuable."

John Creswell, University of Nebraska–Lincoln

Watch video interviews with leading methodologists

Explore the **Methods Map** to discover links between methods

Search a custom-designed taxonomy with more than 1,400 qualitative, quantitative, and mixed methods terms

Uncover more than 120,000 pages of book, journal, and reference content to support your learning

Find out more at
www.sageresearchmethods.com